AN EVALUATION OF THE SAMPLING DESIGN ADOPTED BY THE CENTRAL MARINE FISHERIES RESEARCH INSTITUTE FOR ESTIMATING MARINE FISH PRODUCTION OF INDIA

M. KRISHNAN KUTTY
Regional Centre, National Institute of Oceanography, Cochin-18

A. K. KESAVAN NAIR AND S. Z. QASIM
Central Marine Fisheries Research Institute, Cochin-18

ABSTRACT

The sampling design used by the Central Marine Fisheries Research Institute, for collecting the catch data of sea fish in India is based on the procedure of stratification in space and time. The coastline is divided into different zones and one zone and a cluster of 10 days are taken as the sizes of a space-time stratum. Generally one survey staff is allotted to each zone. Certain limitations in the sampling system seem at present inevitable because of the fishing conditions under which the survey work is being carried out.

The need to consider fluctuations in availability and abundance of fish populations in designing a sample survey is indicated. Fish landings fluctuate widely in space and time. Therefore, a single survey, as is customary in various other surveys, does not give reliable estimates of the strata variances which are essential for the optimum allocation of the sampling personnel. Data for several years are necessary for this allocation. Stratum size should be made sufficiently large so as to make the strata variances more or less the same every year. The increase in the sampling fraction and the distribution of the sampling personnel on the basis of optimum allocation improve the accuracy of the estimates.

Optimum allocation with respect to space-time stratification involves certain amount of operational inconvenience. A state-wise allocation of the total survey staff based on the average yield, regardless of seasonal variations, is suggested as an approximation of optimum allocation. This appears to be a more efficient method than the existing system of giving equal weightage to each zone within a state and ignoring the differences in the state-wise landings. The west coast seems very poorly sampled at present as compared to the east coast. If at least one person is to be allotted to each of the existing zones, the near optimum allocation requires a total strength of 108 field staff, as against the existing 57. Most of the additional staff seems to be needed for the west coast.

INTRODUCTION

India is among the few nations bordering the Indian Ocean which has adopted a sampling system based on the theory of sampling for collecting the fish catch statistics (Banerji, 1971). The sampling design was developed by the CMFRI more than a decade ago and the same with consistent modifications is being followed since then. However, the estimates of fish production made by the Institute on a regional basis and those determined independently by some of the Fisheries Departments of maritime states of India are often not in agreement. For this reason, and since accurate catch statistics are important in stock-assessment studies, it has become necessary to examine the efficiency of the existing sampling system in detail. In this paper we have attempted to answer : 1) to what extent the catch statistics at the all India and states levels are accurate, 2) whether any improvement in the sampling procedure is possible and 3) whether the present sampling fraction, which depends on the number of survey staff, should be increased.

The present study, although preliminary, has clearly shown the importance of a more extensive investigation of the sampling problem, despite the time and effort that may be required for the collection and analysis of data.

Sampling Difficulties

In India, unlike in many advanced nations, fishing is still not a fully organised industry. It basically consists of a large number of persons scattered all along the coastline, who go for fishing by using indigenous craft and gear which consist of a boat and net. There are approximately 1,200 landing centres distributed along the coastline. Therefore, a complete enumeration of the catch and effort, of the type available in advanced fishing nations, is not possible. This makes it necessary to develop a suitable sampling system, which is different from those adopted by the advanced nations, and yet suitable to our needs and conditions. In many ways, the conditions prevalent in India are probably common to many other developing countries.

The data maintained by the Institute, are collected by a team of staff specially employed for the purpose. The landing centres being too many and distributed over an extensive coastline of 5650 kilometres, a large number of data collectors (enumerators) is required, if a large proportion of the landing centres is to be adequately covered. Fishing is carried out in varying intensities throughout the day and sometimes even at night. The landing centres are often located in remote areas which are not easily accessible by any public transport.

These are some of the common features which must be taken into account while choosing a particular sampling procedure.

The Sampling System

Bal and Banerji (1951) have given an account of the effort made earlier by the Institute in developing a planned survey. Several pilot surveys were conducted in different regions of the country by the ICAR between 1950-51 and 1954-55 (Banerji and Chakraborty, 1972). Sukhatme, Panse and Sastry (1958) conducted another pilot survey of the Malabar coast. The present system of stratified multi-stage sampling was developed from these pilot surveys. In large scale sample surveys, sub-sampling becomes essential and stratification—which is the division of a population into such sub-populations within which variability between the units is less than when it is not divided—makes sampling from a highly variable population efficient.

The CMFRI conducted frame surveys during the years 1947-48, 1956-57 and 1961-62. From these surveys, some of the essential details such as the total number of landing centres connected with the sampling system became known.

On the basis of the total number of landing centres, the entire coastline was divided into 56 geographically contiguous zones. The total number of landing centres included in each zone ranged 11-30, depending upon the fishing practices and operational convenience (Table 1). At present the landing centres

TABLE 1. *Number of zones and fish landing centres in the costal states of India*

State	Number of zones	Total number of marine fish landing centres
West Bengal & Orissa	3	45
Andhra	9	253
Tamil Nadu	15	338
Pondicherry	2	23
Kerala	9	215
Mysore	6	98
Maharashtra	8	179
Gujarat	4	79*
Total	56	1230

* The number does not include Kutch area

are controlled by 40 zonal headquarters or survey centres shown in Fig. 1. As an example, the 9 zones of Kerala region together with the number of landing centres have been indicated in Fig. 2.

Each survey centre is housed in 1-2 room rented apartment. Beside the essential furniture, each centre is provided with necessary literature connected with the identification of fish, Institute's important publications, a reference collection of local fishes, crustaceans and molluscs, field note-books and registers. At places where the survey centres are located at the Institute's larger branches (Sub-stations and Units etc., see Fig. 1), facilities for research including a library and a museum plus the assistance of the research staff are available to the survey staff (enumerators).

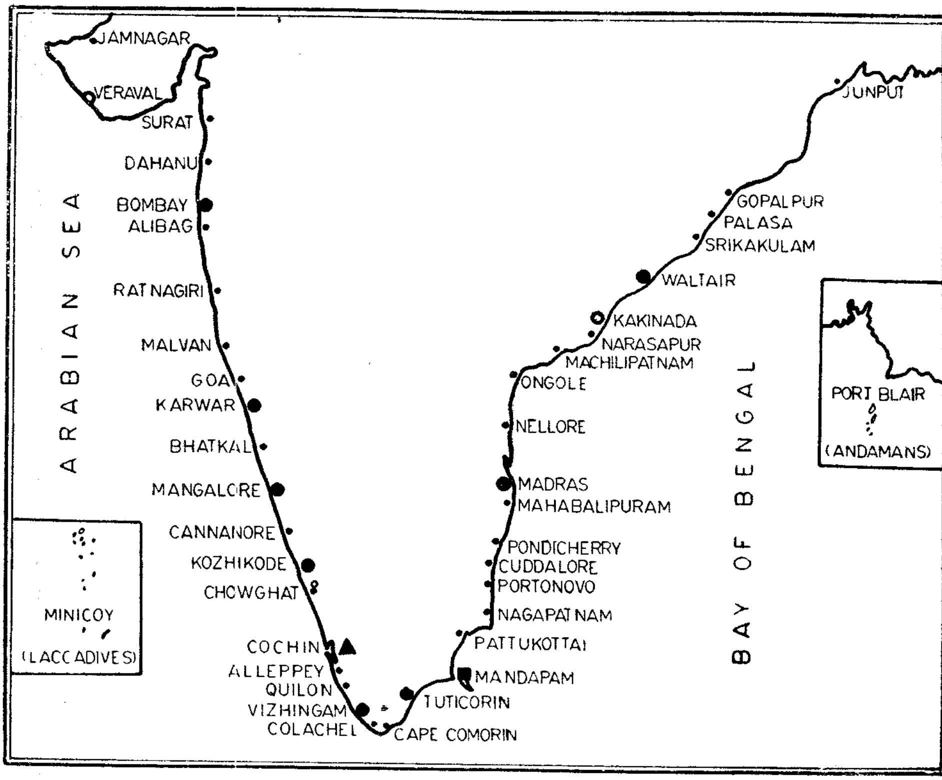

FIG. 1 Map showing the Survey Centres of Central Marine Fisheries Research Institute, along the coastline. Some of the survey centres are housed in larger establishments of the Institute.

▲, Headquarters ; □, Regional Centre ; ●, Sub-stations ; ○, Units, • , Survey Centres

The survey staff, immediately after recruitment, undergoes a training course which lasts 10-12 weeks, and is then posted to the survey centres. Towards the end of each month, the survey staff receives by post the programme of work for the following month, which includes details such as day and time of observations and landing centres to be worked etc. The programme is carefully designed and is sent from Cochin (Institute's Headquarter). Surprise inspections are carried out at frequent intervals by the supervisory staff of the Institute and the enumerators are checked while at work in the field and their field note-books and diaries are scrutinized and initialled.

For observation purposes, a month is divided into 3 clusters, each of 10 days (not counting the last day in months with 31 days). From each cluster, 6 consecutive days are selected according to the following conventions : from the first five days of the first month in a year, a day is selected randomly, which together with the next 5 consecutive days (6 days in all), form the first group.

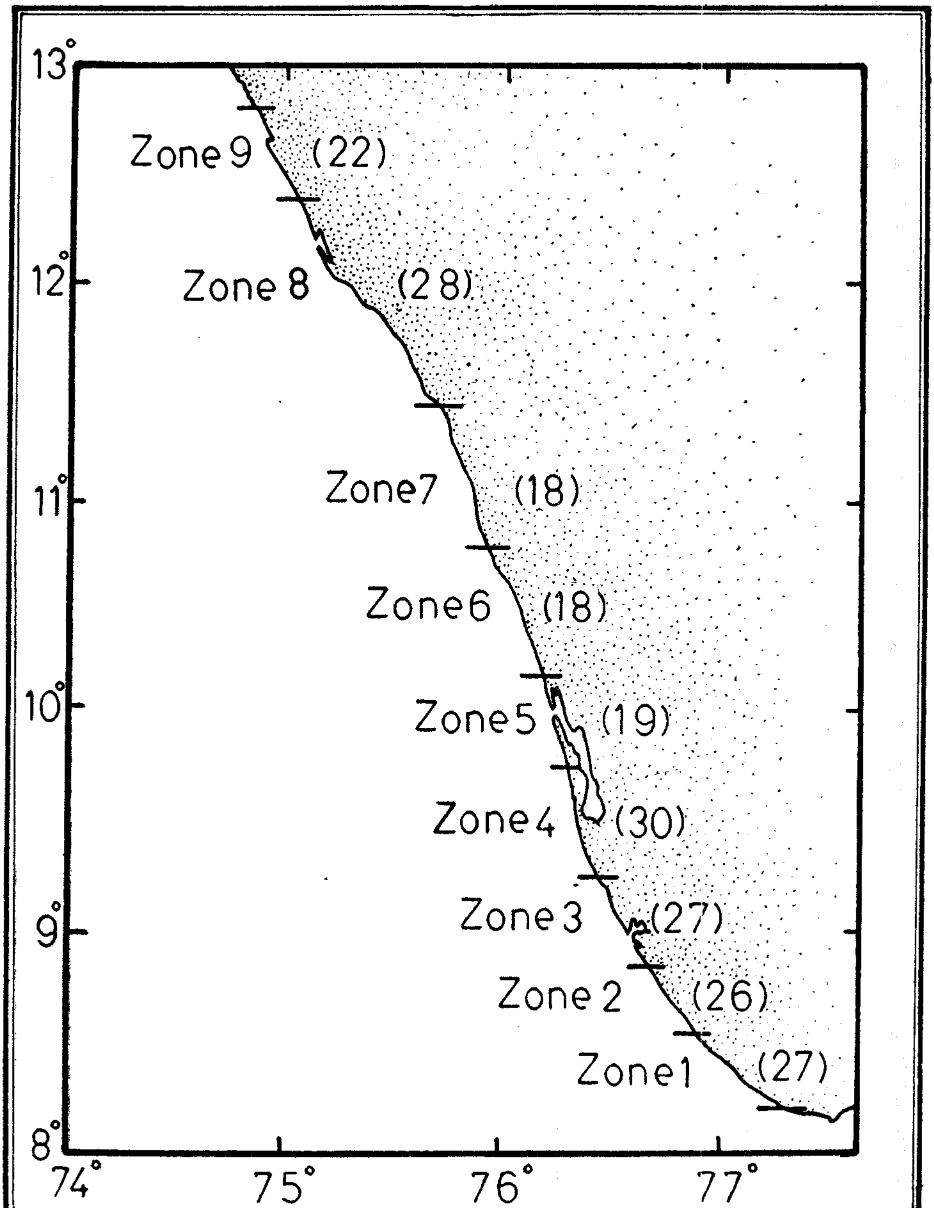

FIG. 2 Map showing the coastline of Kerala divided into 9 zones. The number of landing Centres included in each zone have been shown in parenthesis.

The next 6 days from the other groups are so selected that a 10 day gap falls between the day of the first observation and the next series. For example, if for a zone, the observations have started from the 4th of a month and have been continued until the 9th, the next series will begin from the 14th of that month and subsequent series on the 24th.

Three centres are randomly selected for the 6 day observations and each selected centre is observed for two days (once in the afternoon and the next day in the morning), for a 6-hour duration each day and from these, the landings for a day (12-hours) are estimated. The night landings obtained by enquiry on the second day are added to this figure so as to arrive at the landings for one day (24-hours). The sampling error involved in estimating the landings in this manner is ignored. Thus in a 10-day period, 3 days landings at 3 centres are determined and in a month, 9 landing-centre-days (3 × 3) are sampled from which the total monthly catch is estimated for each zone. The catch for a particular state is then computed by adding the zonal values. Similarly the catch for the entire country is obtained by pooling the estimates of all states.

This is the general pattern of the system employed. Minor deviations in the procedure have been introduced from one state to the other to suit the operational convenience and the nature of landings. Thus, for instance, only two landing centres are sampled in some of the east coast states during a cluster of 10 days to arrive at the landings for a centre-2-days-group. An outline of the details followed in different states has been given by Banerji and Chakraborty (1972).

The above design is characterized by a two way stratification; that is, stratification in space and time. A zone is taken as a space stratum and a cluster of 10 days becomes a time stratum for sampling. For computational purposes, a calender month is taken as the time stratum. When the fish landings are found to vary considerably between the landing centres of a zone, especially in different seasons, a zone is further stratified (sub-divided) by grouping the good landing centres to form a separate stratum (group). Only one field staff is generally provided for a particular zone.

For sub-sampling, a landing-centre-day is selected as the first stage unit (fsu). Thus if there are 20 landing centres in a zone, there will be 20×30 first stage units for a month in that zone. In certain states, especially on the east coast, a centre-2-days-group forms the primary unit. Here the 12 hours-day landings are obtained from actual observation for the entire day or from 6 hours of observation in two periods of 3 hours each. Thus from the 2-day observations, the quantities landed during the days (12 + 12 hours) are obtained. The quantities landed during the two nights are obtained by enquiry and from these the landings for a centre-2-days-group are determined.

Since it is not easy to observe the catches of all the boats belonging to a particular landing centre, sampling of the boats becomes essential. Thus a boat is selected as the second stage unit (ssu). When the total number of boats landed is 10 or less, the total landings from all the boats are enumerated for catch composition and for other particulars. But when the total number of boats exceeds 10, the following percentages are sampled.

No. of boats landed	Percentage sampled
11-20	50
21-50	20
51 and above	10

From the boats, the catches are usually removed in baskets of standard size. The weight of fish contained in these baskets being known, the weight of fish in each boat under observation is obtained.

Estimation of Total Catch

As noted above the figures for landing during the day and night are put together to give the centre-day-catch. From these the monthly zonal landings are obtained as :

$$\hat{Y}_{ijk} = \frac{N_{ijk}}{n} \sum_{l=1}^{n} y_{ijkl} \qquad \ldots\ldots(1)$$

where $\hat{Y}_{ijk}$ is the estimated catch for the k^{th} month for the j^{th} zone in the i^{th} state, N_{ijk} is the number of landing centre-days for the k^{th} month for the respective zone, n is the corresponding number of centre-days actually sampled and y_{ijkl} is the estimated yield for the l^{th} landing centre-day sampled from the k^{th} month for the j^{th} zone in the i^{th} state.

By pooling the zonal estimates, the monthly fish landings of the state and for the country as a whole are obtained. The annual catch is similarly estimated by summing the monthly landings.

Errors in Estimate

Sukhatme *et al.* (1958) have shown that the correlation between the number of boats landing per hour and the average catch per boat for the same hour is small. They have further shown that the coefficient of variation of the catch per boat is also small. This is mainly because of the indigenous method of fishing in which the fishing power, the gears employed, the time spent on fishing and the area fished remain nearly the same. Hence the error introduced by sampling the ssu remains minimum and can therefore be ignored. Panse and Sastry (1960) have shown that this error is also small for the Egyptian fishery.

Nevertheless, the estimate of the centre-day-landings may be subjected to errors from the mode of enumeration of the selected centre-day. The night-landings obtained by enquiry during the following morning (next day) are only approximations. Moreover, some error is likely to occur even in the estimates of the day-landings, for these are not completely covered. The systematic selection of the clusters of hours is based on the analysis of the efficiency of the different methods of sampling a day conducted by Sukhatme *et al.* (1958). But all their recommendations could not be followed because of practical difficulties involved in their implementation. These difficulties have been discussed by the authors in some detail, who have also shown that systematic sampling is more efficient than simple random sampling when the sampling unit is either an hour or a cluster of hours (Sukhatme *et al* 1958). The efficiency, however, was found to decrease with the size of the cluster.

The magnitude of error in the estimates of centre-day-landings, under different procedures of the day sampling, could not be examined in detail. However, the selected centre-day is enumerated completely to the extent possible by the field staff. On this basis, the centre-day-catch, as suggested by Sukhatme *et al.* is assumed to be determined without error, so that the estimates of variances of the yield can be obtained from the following equations :

The estimate of the variance of $\hat{Y}_{ijk}$ is

$$v(\hat{Y}_{ijk}) = \frac{N^2_{ijk}}{n} v_{ijk} \qquad \ldots\ldots (2)$$

where v_{ijk} is an estimate of the stratum variance which is given by

$$v_{ijk} = \frac{\sum_{l=1}^{n} y^2_{ijkl} - \frac{\left(\sum_{l=1}^{n} y_{ijkl}\right)^2}{n}}{n-1} \qquad \ldots\ldots (3)$$

For estimating v_{ijk}, the finite population correction factor (fpc) is ignored, as it is approximately equal to unity. By pooling the zonal estimates of the variances, that is, within and between the states, both monthly and yearly estimates for the states and for the country as a whole can be obtained. Thus the variances of the estimated annual yields $\hat{Y}_{i..}$ and $\hat{Y}_{...}$ at the state and all India levels respectively are :-

$$v(\hat{Y}_{i..}) = \sum_{j=1}^{z_i} \sum_{k=1}^{12} v(\hat{Y}_{ijk}) \qquad \ldots\ldots (4)$$

and

$$v(\hat{Y}_{...}) = \sum_{i=1}^{v} \sum_{j=1}^{z_i} \sum_{k=1}^{12} v(\hat{Y}_{ijk}) \qquad \ldots\ldots (5)$$

where z_i is the number of zones in the i^{th} state and v is the number of costal states.

The confidence interval at the 95% level for the annual national yield is

$$\hat{Y}_{...} \pm 2\sqrt{v(\hat{Y}_{...})}$$

which for 1966 is approximately ± 12% of the estimated all India landings. For the same year the confidence interval for the different states ranges from ± 11% to ± 43% of the estimated landings in different states.

LIMITATIONS IN THE SAMPLING SYSTEM

In the sampling system, the introduction of space-time stratification becomes necessary as the fish population varies both in space and time. Stratification makes a heterogeneous population more uniform with respect to eaeh stratum, and the present sampling system, by virtue of stratification, reduces the variance significantly. However, the distribution of the sample size to each stratum (zone) on the basis of one enumerator for a zone does not seem to be the best arrangement because of the wide difference in regional landings. The annual landings on the east coast is ~ 25% and on the west coast ~ 75% of the total. Variance generally increases with the size of the landings and therefore, if the error in the estimate is to be reduced, the number of sampling personnel allotted to the west coast should be increased. It is therefore necessary to examine whether the sampling personnel can be distributed on the basis of 'optimum allocation' which takes into account the variability in regional differences. This will increase the accuracy of the estimates of all India and state landings.

Apart from this, the other aspects which require careful examination are: (a) the sampling fraction and (b) the efficiency of sub-sampling the fsu (centre-day). Even when the centre-day-landing is assumed to be determined without error, the present sampling fraction is very low and ranges from 1.5 to 3% in different states. As pointed out earlier, because of several difficulties, the sampling fraction cannot readily be increased to the desired level. However, the question of increasing the sampling fraction deserves serious consideration, as accurate catch statistics are necessary for stock assessment. With the existing sampling fraction of 1.5 to 3%, the all India landings lie within ± 12% of the estimated landings. At the states level the confidence interval for the landings is much wider.

OPTIMUM ALLOCATION IN RELATION TO SPACE-TIME STRATIFICATION

When the allocation of the field staff is done in order to minimise the error of the estimate for a fixed cost or to minimise the cost for a given precision, it is termed as optimum allocation. If it is possible to work out an allocation on this basis (from one of these principles), the net result of the survey would be immensely advantageous. Assuming that the average cost of enumerating a landing centre in each zone is the same, the optimum allocation of the number of sampling personnel to the different zones will depend upon the size of the stratum or zone (i.e. the number of landing-centre-days in the zone) and the zonal variances which are not known. Sukhatme (1953) has suggested a method of estimating the stratum variance from a preliminary sample. The minimum size of this sample is given by

$$n' \geq \frac{\left(\sum_{i=1}^{k} p_i S_i \right)^2 - \sum_{i=1}^{k} p_i^2 S_i^2}{2\sum p_i (S_i - \overline{S_w})^2} \qquad \ldots\ldots (6)$$

where p_i is the relative size of the i^{th} stratum, S_i is the standard deviation of the i^{th} stratum, k is the number of strata and $\overline{S_w}$ is $\sum_{i=1}^{k} p_i S_i$. The stratum variance is then estimated using a sample size of n' for each stratum.

However, it is necessary to examine whether in fish population studies, a reliable estimate of the stratum variance can be obtained and which can be used for determining the optimum allocation. Table 2 gives the estimates of the strata variances (v_{ijk}) for the different zones of Kerala for 1967 and 1968. As can be seen from the table, the strata variances are neither stable nor do they show a similar trend year to year.

Several factors are responsible for making the present estimates of strata variances unsatisfactory for their use in determining the optimum allocation. Marine fish population, unlike terrestrial populations, have several unique features and these must be taken into account in designing a sampling programme. For example, in agricultural surveys where the population is more or less under human control, the variability between the years is generally not significant and can therefore be ignored. In marine fish populations, on the other hand, several extraneous factors operate and the populations exhibit wide annual fluctuations, especially if they are of pelagic species. Changes in their migratory pattern largely control their availability in a particular zone or stratum. Nearly 35% of the annual all India landings is contributed by the two pelagic stocks, namely the oil sardine and mackerel. These two species are largely confined to the south-west coast of India. For these reasons—as the annual changes in availability and abundance have to be accounted for—the estimates of strata variances based on one year's data will not be reliable. Many years data are therefore necessary for estimating the strata variances.

These features also make it necessary to examine the size of the space-time stratum required for obtaining a stable estimate of the stratum variance. The existing estimates are based on a stratum, the size of which with respect to space is a zone and with respect to time, a month. When the space-time stratum is as small as this, the stratum variance over the years will fluctuate widely. Hence allocation using the estimates of variances with respect to such small strata will not be efficient even when many years data are available.

TABLE 2. *Estimated variance (ton²) of the yields of landing-centre-days for the 9 zones of Kerala*

Zone	January		February		March		April		May		June	
	1967	1968	1967	1968	1967	1968	1967	1968	1967	1968	1967	1968
I	0.55	7.44	2.59	9.01	1.02	0.48	16.25	6.22	No data	16.42	2.95	18.00
II	9.01	0.35	23.92	0.34	1.93	3.69	2.91	3.22	4.09	7.07	38.58	5.10
III	48.88	201.23	100.62	75.83	16.12	5.06	7.55	5.56	3.10	0.24	0.46	0.19
IV	21.37	43.67	32.56	97.59	18.84	9.93	18.40	3.77	0.40	0.68	60.13	409.95
V	53.88	444.95	61.29	1,396.32	957.80	92.70	5.58	21.35	15.99	26.21	31.03	1.06
VI	966.39	288.44	1,064.32	87.72	477.57	75.88	251.88	21.23	73.37	7.80	86.16	4.68
VII	497.92	95.47	203.94	74.84	0.94	12.22	63.54	No data	240.54	93.91	29.63	17.96
VIII	15.29	1,733.94	1,211.45	301.77	126.20	184.55	80.62	38.20	91.45	279.35	8.63	No data
IX	4,769.31	926.24	1,261.66	2.97	365.79	126.05	2.81	105.76	4.21	1.34	0.00(5)	No landings

zone	July		August		September		October		November		December	
	1967	1968	1967	1968	1967	1968	1967	1968	1967	1968	1967	1968
I	10.17	0.12	14.66	3.01	26.84	13.29	1,662.15	14.67	4.22	2.64	3.20	2.37
II	0.41	1.64	14.36	1.32	680.13	95.91	143.48	14.97	23.03	3.47	0.83	16.76
III	72.65	2.84	6.37	9.14	53.82	21.08	859.31	16.31	98.12	40.97	12.06	293.46
IV	3,338.89	3,990.97	0.45	161.17	1.35	40.82	149.13	35.04	459.97	175.39	18.67	36.73
V	0.03	21.99	315.72	110.50	2,212.11	1,477.69	1,696.94	5,109.31	347.92	5,478.43	2,818.19	2,818.19
VI	48.99	45.23	36.09	18.77	127.54	34.81	196.28	58.91	103.94	169.98	77.06	128.10
VII	1.22	0.00(1)	32.02	108.02	35.67	3,510.43	162.37	1,174.45	735.69	3,441.79	1,226.38	585.94
VIII	270.77	270.77	7,433.90	1,954.30	3,032.80	15,134.01	18,985.54	8,305.80	8,603.07	4,021.64	10,487.96	1.75
IX	3.86	0.04	233.83	788.25	122.78	2,850.83	1,344.90	1,035.15	1,150.11	1,428.11	34,086.41	1,248.27

Table 3 gives the state-wise percentage errors of the estimated annual landings for five years. At the state level, the percentage error appears to be

TABLE 3. *Percentage errors of the estimated annual yields for some states during the years 1966-1970. Blanks indicate that for some years in respect of some states, percentage errors could not be calculated, as the data were not in a form to allow computation with ease.*

Costal states	1966	1967	1968	1969	1970
Andhra	6.44	8.89	8.22	11.50	7.58
Tamil Nadu	5.34	...	...	...	...
Kerala	8.20	9.58	6.89	7.67	...
Mysore	17.80	27.12	...	...	18.80
Maharashtra	21.26	18.25	10.00	9.92	7.46

reasonably uniform. Hence the state-wise variance is expected to be more stable than the zonal estimates given in Table 2. Therefore, for optimum allocation, a state may be taken as a stratum with respect to space. Size of the space stratum may also be determined on the basis of regional differences in fish landings, but we have not examined the increase in accuracy by adopting this procedure.

The total landings (all fishes inclusive) of a state will not vary very much from year to year, but the landings of individual species, particularly of the oil sardine and mackerel, may exhibit wide variations within the year and between the years. An important aspect of collecting catch statistics, in addition to the estimation of total landings, is the assessment of the commerically important species. The sampling programme should, therefore, be so designed that as far as possible accurate data with respect to each economically important species are obtained. Separate sampling for individual species is not advantageous as fishing is not normally done for any one particular species at a time. However, by increasing the sample size during the peak season, reliable estimates can be obtained. The fishing season for the oil sardine and mackerel generally lasts for 8-9 months. It has a lean period of about 3 to 4 months. The year can therefore be conveniently divided into distinct periods, each of 3-4 months, depending upon the fishing intensity and each period can be treated as a time stratum for the west coast. This period should be fixed in such a way that the landing remains as far as possible, homogeneous within each time stratum.

Because there is no major seasonal fishery along the east coast, one whole year may be treated as the time stratum for the east coast, so that the year to year variations in the stratum variance may remain small. A state and a period of 3-4 months along the west coast, can be taken as the sizes of the space-time strata for the purpose of optimum allocation.

Allocation of the survey staff to the enlarged strata based on the variances calculated from the data of several years becomes very meaningful and this can be worked out from the general formula. But if the general formula for determining optimum allocation (see e.g. Cochran, 1953; Sukhatme, 1953; Murthy, 1967) is used, in such cases where space-time stratification is involved, the sample size during the peak season may become so large that the available sampling personnel may prove to be totally insufficient. Similarly, during the lean season, the personnel will not have full work-load. Therefore, the following modifications in the general procedure for the determination of optimum allocation become necessary.

The number of enumerators for the west coast is first determined for the region as a whole. Then the enumerators are distributed between the different states (strata with respect to space) separately for each time strip. One year being the stratum for the east coast, the allocation of the survey staff for the different states can be easily done without considering the season. The number of personnel (P_w) for the west coast can be obtained from:

$$P_w = P \frac{N_w \sqrt{v_w}}{N_w \sqrt{v_w} + \sum_{i=1}^{n_{se}} N_{ei} \sqrt{v_{ei}}} \qquad \ldots\ldots (7)$$

Where P is the total number of available personnel, N_w is the number of landing centre-days for the west coast as a whole for the year, N_{ei} is the number of landing centre-days for the i^{th} stratum in the east coast for the year, n_{se} is the number of strata in the east coast, v_w is the estimated variance for the west coast and v_{ei} is the estimated variance for the i^{th} stratum in the east coast. The number of personnel for the i^{th} stratum in the east coast is given by

$$P_{ei} = P \frac{N_{ei} \sqrt{v_{ei}}}{N_w \sqrt{v_w} + \sum_{i=1}^{n_{se}} N_{ei} \sqrt{v_{ei}}} \qquad \ldots\ldots (8)$$

Similarly the personnel or enumerators, P_{ij} for the i^{th} space stratum of the west coast for the j^{th} time strip is obtained from:

$$P_{ij} = P_w \frac{N_{ij} \sqrt{v_{ij}}}{\sum_{i=1}^{n_{sw}} N_{ij} \sqrt{v_{ij}}} \qquad \ldots\ldots (9)$$

where N_{ij} is the number of landing centre-days in the i^{th} space stratum for the j^{th} time stratum, n_{sw} is the number of strata with respect to space in the west coast and v_{ij} is the estimate of the variance for the i^{th} space stratum for the j^{th} time stratum.

For estimating the yields and the variances of the strata, it may not be desirable to adopt the usual simple random sampling procedure partly because of operational incovenience and partly because it is not the best method when the stratum size is fairly large and when there are local differences in the fishing practice and in the size of the gear employed. For the latter reason, even restricted random sampling may have certain difficulties, as the stratum yields are to be estimated gear-wise. Hence the best approach seems to be to sub-divide a stratum into smaller regions or sub-strata. The existing zones can be treated as sub-strata. The yield and variances are first estimated with respect to these sub-strata and then an estimate of the stratum variance (v) for the purpose of optimum allocation is obtained from :

$$v = \sum_{i=1}^{k} \frac{N_i}{N} v_i + \frac{N}{N-1} \left\{ \sum_{i=1}^{k} \frac{N_i}{N} (\bar{y}_{ni} - \bar{y}_w)^2 - \sum_{i=1}^{k} \frac{N_i}{N} \left(1 - \frac{N_i}{N} \right) \frac{v_i}{n_i} \right\} \quad \ldots(10)$$

where N_i is the number of landing-centre-days in the i^{th} sub-stratum, N is the number of landing-centre-days in the stratum of allocation, n_i is the number of landing-centre-days actually sampled from the i^{th} sub-stratum, $\bar{y}_{ni}$ is the estimated mean landing centre-day yield for the i^{th} sub-stratum, v_i is the estimated variance for the i^{th} sub-stratum and

$$\bar{y}_w = \sum_{i=1}^{k} \frac{N_i}{N} \bar{y}_{ni} \quad \ldots\ldots (11)$$

If such estimates of v are available for r years, an unbiased estimate of the stratum variance $\left(\hat{v} \right)$ can be obtained from

$$\hat{v} = \frac{n^{(1)}v^{(1)} + n^{(2)}v^{(2)} + \ldots + n^{(r)}v^{(r)}}{n^{(1)} + n^{(2)} + \ldots + n^{(r)}} \quad \ldots\ldots (12)$$

where $n^{(1)}, n^{(2)}, \ldots, n^{(r)}$ and $v^{(1)}, v^{(2)}, \ldots, v^{(r)}$ are the respective total stratum sample size and the stratum variances for the r years. By treating the entire west coast as one stratum, the equations (10)-(12) can be used for estimating its variance.

For distributing sample size on the basis of optimum allocation, Hansen *et al.* (1953) have suggested that the average stratum production per unit can be used as an approximation to the stratum standard deviation (square root of the stratum variance) if a similar trend in the stratum production is maintained between the years. This suggestion is based on the assumption that the average production per unit is proportional to the standard deviation. Table 4 gives the state-wise marine fish production for a period of 10 years. It is clear from the table that the general trend in the fish landings of a state (taken as a stratum) does not change over wide limits. Generally, variance increases with the size of the landings and hence this approximation to the optimum allocation should remain valid for fish populations. Therefore, the equations (7), (8) and (9) may be used after replacing the estimated standard deviation of the stratum by the corresponding mean yield per landing-centre-day, especially when a reliable estimate of the stratum variance is not available. The allocation will now be proportional to the size of the landings.

The optimum allocation based either on the stratum variance or on the landings may prove to be inconvenient because it will introduce frequent transfer (once for each time stratum) of the field staff working on the west coast from one

TABLE 4. *Annual state-wise marine fish landings in India from 1962 to 1971. All figures are in thousand tonnes*

State	Year										Mean landings
	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971*	(Geometric mean)
West Bengal & Orissa	7·9	11·1	10·6	12·2	10·0	19·3	30·7	22·9	31·4	27·3	16·3
Andhra	60·0	64·6	71·7	76·5	80·1	76·1	77·4	77·5	74·5	83·3	73·8
Tamil Nadu (Including Pondicherry)	111·4	109·5	131·3	106·0	147·5	160·5	162·5	162·5	166·1	170·7	140·6
Kerala	191·4	202·4	318·0	339·2	346·7	364·1	345·3	294·8	392·9	445·6	314·3
Mysore	43·9	36·5	104·2	68·5	65·6	49·2	87·8	75·8	116·9	104·0	70·4
Goa	Not Available			17·2	24·6	12·5	18·9	27·5	20·7	21·0	19·8
Maharashtra	123·7	121·3	130·6	131·9	134·3	133·3	123·9	168·7	192·4	216·3	144·7
Gujarat	97·7	101·9	92·9	80·6	80·3	75·6	86·6	82·2	89·0	84·9	86·8

* Provisional

state to the other. If the optimum allocation cannot be implemented in this way, the state-wise allocation on the west coast should be done on an annual basis ignoring the seasonal differences. This is to be done at the loss of some accuracy. The distribution of personnel in respect of different states—the existing and that based on the mean annual landings—has been given in Table 5. It is evident

TABLE 5. *State-wise allocation of the field staff for the collection of catch data when mean annual landings is used in weight*

Coastal States	Existing distribution of 57 field staff	Weighted distribution or near optimum allocation of 57 personnel	by allowing at least one person in each zone of the state
	WEST COAST		
Gujarat	5	6	11
Maharashtra	8	9	18
Mysore	6	6	11
Kerala	9	21	39
	EAST COAST		
Tamil Nadu including Pondicherry	17	9	17
Andhra	9	5	9
West Bengal & Orissa	3	1	2
Total	57	57	107

from the table that at present some over representation has been given to the east coast and that the Kerala coast has been too much under staffed. The east coast has in the past been devoid of any major fishery of the kind one finds on the west coast. It has also been characterised by less fish production than on the west coast. The fish landings on the east coast are largely composed of such species which have been thought to have less economic importance than those on the west coast. All these factors make it possible to have a fewer number of sampling staff on the east coast as compared to that of the west coast. If 107 field staff (enumerators) are distributed as per allocation indicated in Table 5, all states, except West Bengal and Orissa will get at least one person in each zone. As West Bengal and Orissa get only 2 persons for the 3 zones of these states, a practical suggestion would be to allot 1 more person to these two states so that there would be at least one person for each zone. Hence the total number of sampling personnel reaches 108. Since the existing staff of 57 is too low, the redistribution, as indicated in column (2) of Table 5, does not seem practicable. However, when the number of staff increases, almost all extra hands should go to the west coast. They should be allocated to the different states in proportion to the fish landings.

Allocation Within a Stratum (State)

It would not be advantageous to distribute the survey staff within each stratum (state) on the basis of optimum allocation using zonal variances or zonal landings, as these two are not likely to remain stable. The stratum (state) is to be sub-divided for the operational convenience and for minimising the time spent on travelling and to account for the differences in fishing practices. Therefore, the existing zones and the present allocation of at least one person for a zone seem desirable. Additional staff when available should be distributed within the productive zones.

Improvement in Sampling the Centre-Day

In certain states, sampling of the centre-day is done systematically in clusters of 6 hours and the sampling is completed in two consecutive days. In Kerala, the cluster size is 3 hours and sampling is completed the same day. Since the efficiency decreases with the size of the cluster (Sukhatme *et al.*, 1958), systematic sampling in clusters of 3 hours is expected to be more efficient than when the cluster-size is 6 hours. However, it may be difficult to adopt a uniform cluster-size of 3 hours due to operational difficulties. Wherever this is not possible and the cluster of 6 hours is to be retained, a suitable weightage should be given to the period when the landings are heavy. If this is done, in certain zones the day will have to be further stratified into forenoon and afternoon for estimating the catch. No improvement in the estimation of night-landings seems possible at present, as direct observations during the night are so difficult to make.

Sampling Fraction

The sampling fractions, calculated on the assumption that the estimates of centre-day-yield are without error, vary from 1.5 to 3% in different states. This assumption seems unavoidable because a complete coverage of the centre-day is not practical, and therefore, the figures 1.5 to 3% seem overestimates. For the same reason, the percentage errors in the estimated landings for the states and the country as a whole seem conservative estimates. The percentage error at the state level (Table 3) is much greater than at the all India level ($\sim$6%). To increase the accuracy of the estimates, a substantial increase in the sampling fraction is necesssary, although it is fully realized that a large increase would not be possible economically. Using the estimates of variances, the number of personnel required to reduce the percentage error from the existing level to any desired level can be calculated. But the existing data for several years are not in a form which could be used for obtaining the estimates of variances of the states. Therefore, for the time being the best recommendation seems to be to increase the sampling personnel and their distribution along the lines shown in Table 5. As an alternative, if the sampling personnel of the CMFRI and those of the Fisheries Departments of different states can be brought under a unified programme, the sampling fraction can be considerably increased without any appreciable increase in the overall expenses. In fact this seems another important recommendation, for the Fishery Departments of many states do not collect catch statistics in a systematic fashion. The wide discrepancy often appearing in the estimates of CMFRI and those of the states, e. g., the figures of Gujarat for 1970, cannot be explained simply on the basis of low sampling fraction by the CMFRI. Of the 4 zones in Gujarat, there is only one zone which contributes nearly 68% of the total catch and this zone is fairly heavily sampled. The sampling fraction for the three major centres in this zone, which together contribute nearly one third of Gujarat's total landings, is about 10%.

Conclusion

1. From the foregoing account it is clear that the sampling design followed by CMFRI is based on recommended statistical principles and is scientifically sound.
2. The procedure gives a fairly reliable estimate of the total all India fish landings; and looking at the fishery structure of the country as a whole, the efforts made by the Institute to arrive at species-wise, gear-wise and state-wise figures, are commendable.
3. The estimates of fish production made for the different maritime states, though less accurate than the all India figure, are nevertheless, realistic.
4. Main limitations in the sample survey are the use of a low sampling fraction and the uniformity in the distribution of survey staff which has been determined by ignoring the regional differences in the size of landings.
5. A sudden increase in the sampling fractions does not seem a practical solution because it would be associated with a substantial increase in the cost of extra survey, supervision of the work of the extra field staff, operational difficulties and extra work-load in data processing.
6. Improvements in the estimates are possible by redistributing the existing field staff on the basis of optimum allocation.
7. Optimum allocation with respect to space-time stratification will involve practical difficulties and hence state-wise allocation on the basis of mean annual landings has been recommended. A total of 108 field staff is necessary if a minimum of one field staff is to be provided to each of the existing zones.

Acknowledgement

We thank Dr. N. K. Panikkar, Director, National Institute of Oceanography, for the encouragement he gave to one of us (MKK) to collaborate in this work.

REFERENCES

BAL, D. V. AND S. K. BANERJI. 1951. A Survey of the Sea Fisheries of India. *Proc. Indo-Pacif. Fish. Counc., Sec.* **11**, *176–181.*

BANERJI, S. K. 1971. Fishery Statistics. Food and Agricultural Organisation of the United Nations, Publ. No. 10 FC/DEV/71/5, 15 pp.

BANERJI, S K. AND D. CHAKRABORTY. 1972. On estimation of yield from exploited marine stocks with reference to South East Asia. In : Proceeding, symposium on Living Resources of the Seas around India, *C. M. F. R. I., Spl. Publ., 176-183.*

COCHRAN, W. G. 1953. *Sampling techniques.* John Wiley and Sons, Inc., 330 pp.

HANSEN, M. H., W. N. HURWITZ AND W. G. MADOW. 1953. *Sample Survey methods and theory. Vol. 1: Methods and applications.* John Wiley and sons, Inc., New York, 638 pp.

MURTHY, M. N., 1967. *Sampling theory and methods.* Statistical Publishing Society, Calcutta, India, 684 pp.

PANSE, V. G. AND K. V. R. SASTRY. 1960. Sample surveys for fishery statistics. Technical Report to the Government of the United Arab Republic, Food and Agricultural Organisation of the United Nations, Expanded Technical Assistance Programme Report No. 1247, 176 pp.

SUKHATME, P. V., 1953. *Sampling theory of surveys with applications.* Iowa State College Press, Iowa, U. S. A., 491 pp.

SUKHATME, P. V., V. G. PANSE, AND K. V. R. SASTRY. 1958. Sampling technique for estimating catch of sea fish in India. *Biometrics, 14* : 78–96.