

# YALE PEABODY MUSEUM

P.O. BOX 208118 | NEW HAVEN CT 06520-8118 USA | PEABODY.YALE. EDU

## JOURNAL OF MARINE RESEARCH

The *Journal of Marine Research*, one of the oldest journals in American marine science, published important peer-reviewed original research on a broad array of topics in physical, biological, and chemical oceanography vital to the academic oceanographic community in the long and rich tradition of the Sears Foundation for Marine Research at Yale University.

An archive of all issues from 1937 to 2021 (Volume 1–79) are available through EliScholar, a digital platform for scholarly publishing provided by Yale University Library at <https://elischolar.library.yale.edu/>.

Requests for permission to clear rights for use of this content should be directed to the authors, their estates, or other representatives. The *Journal of Marine Research* has no contact information beyond the affiliations listed in the published articles. We ask that you provide attribution to the *Journal of Marine Research*.

Yale University provides access to these materials for educational and research purposes only. Copyright or other proprietary rights to content contained in this document may be held by individuals or entities other than, or in addition to, Yale University. You are solely responsible for determining the ownership of the copyright, and for obtaining permission for your intended use. Yale University makes no warranty that your distribution, reproduction, or other use of these materials will not infringe the rights of third parties.



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.  
<https://creativecommons.org/licenses/by-nc-sa/4.0/>



# ESTIMATIONS OF ABUNDANCE OF THE EGGS OF THE PACIFIC PILCHARD (*SARDINOPS CAERULEA*) OFF SOUTHERN CALIFORNIA DURING 1940 AND 1941<sup>1</sup>

BY

OSCAR E. SETTE AND ELBERT H. AHLSTROM

*Fish and Wildlife Service  
United States Department of the Interior*

## INTRODUCTION

One of the more important items of information required for an understanding of the dynamics of population numbers in marine fish stocks is the rate of replenishment of the stock through its reproductive process. It has been established for the Pacific pilchard, or California sardine (*Sardinops caerulea*), that some spawning seasons give rise to much larger year-classes than others (Clark, 1939), as measured at the time the year-class reaches commercial age. But it remains to be determined whether this is due to inequalities in the reproductive process or to variations in the survival rate through any one of the several stages of existence between egg and commercial age.

A requirement for this problem in its entirety is the determination of numbers in each year-class through the several stages: egg, larva, post larva, and juvenile. The first two stages, and perhaps a large portion of the third, are planktonic, and for these stages the problem becomes one in quantitative plankton research. This report is concerned with the enumeration of the eggs in a certain area during two spawning seasons. The eggs, being nonmotile, are perhaps the simplest of the several stages to deal with, yet even their enumeration involves a number of complexities and untested assumptions. The results reported here form a basis for further advances and can in no sense be accepted as final determinations.

The data on which this report is based have been published by Ahlstrom (1948). The hydrographic data for the cruises in which the plankton was collected are given by Sverdrup and Staff (1944, 1947). The sea work was done in collaboration with the Scripps Institution of Oceanography, on its research vessel E. W. SCRIPPS.<sup>2</sup>

<sup>1</sup> Published by permission of the Director, U. S. Fish and Wildlife Service.

<sup>2</sup> The authors gratefully acknowledge the aid given by Milner B. Schaefer in the development of the mathematical treatment.

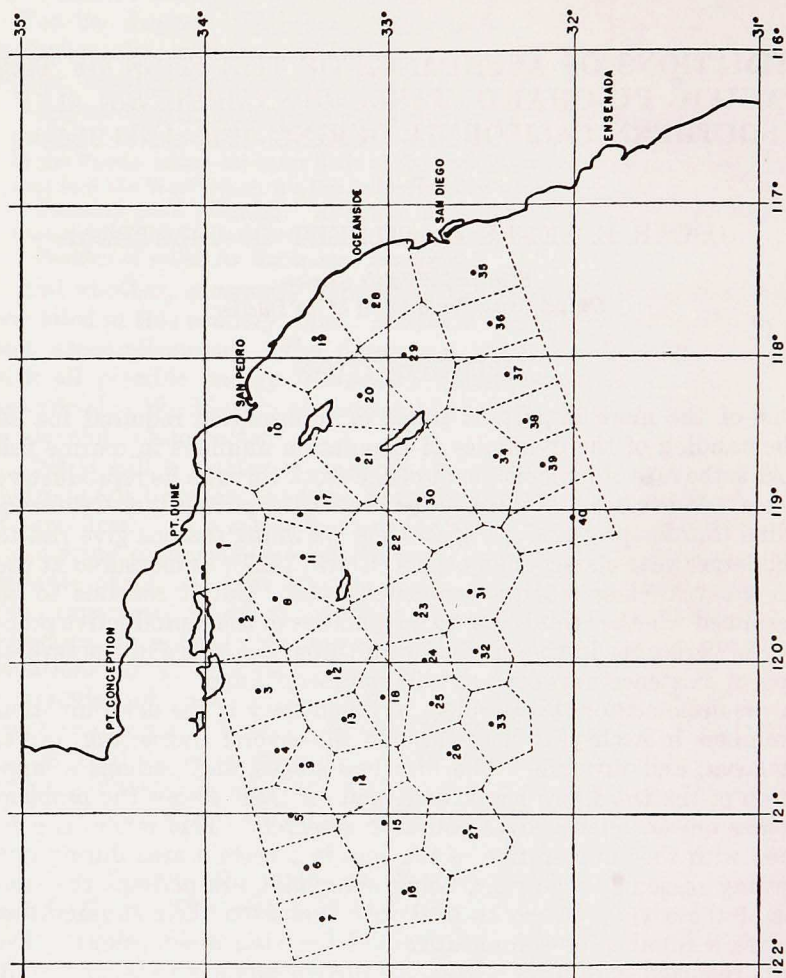


Figure 1. Chart of cruise XIII, showing the station pattern followed during the 1940 survey and the polygons used for area weighting.

The spawning range of the pilchard, though not definitely known, may extend from southern Lower California to at least as far north as off Central Oregon, and in some places as far seaward as 300 miles. Since this entire range could not be surveyed by the single vessel available, the surveys of 1940 and 1941 were confined to a limited area off southern California which was known to include important amounts of spawning (Scofield, 1934).

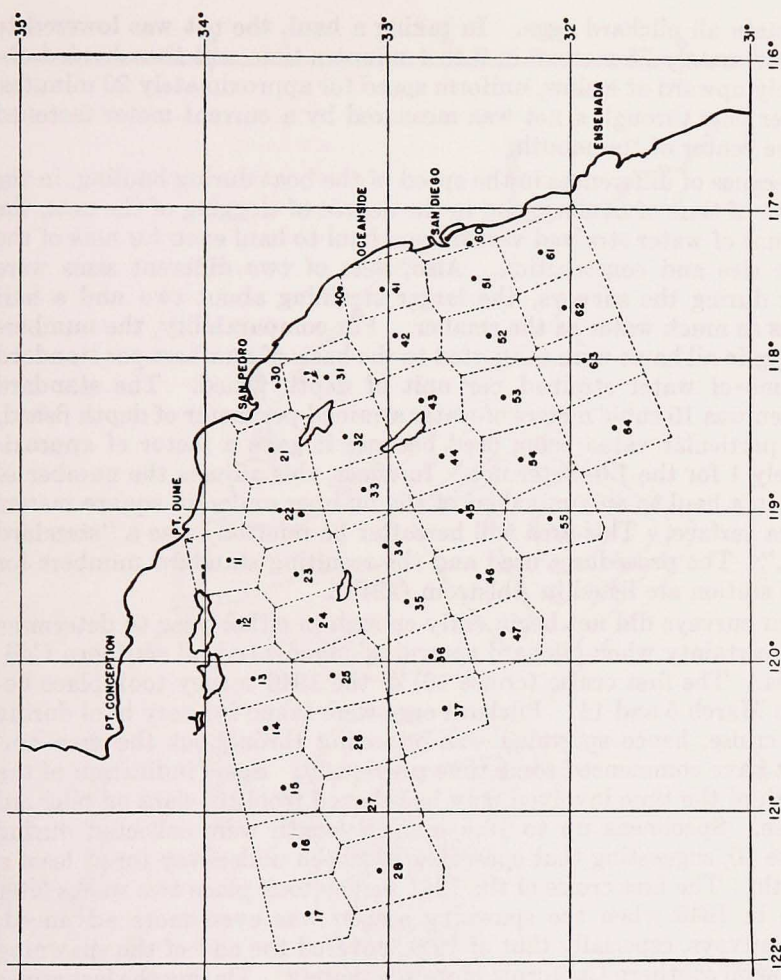


Figure 2. Chart showing the station pattern followed during the 1941 survey and the polygons used for area weighting.

A pattern of stations was established in this area, as shown in Figs. 1 and 2, for 1940 and 1941, respectively. There were six cruises each season, giving a mean interval between successive observations at a station of about 17 days during 1940 and about 25 days during 1941.

The plankton nets employed during 1940 were 1.6 meters in diameter at the mouth, while during 1941 they were 1.0 meter in diameter. The nets were constructed of either No. 24xxx grit gauze or of cotton scrim of similar mesh size. Both materials were of fine enough mesh

to retain all pilchard eggs. In taking a haul, the net was lowered to approximately 75 meters<sup>3</sup> in 2 to 4 minutes time and then hauled obliquely upward at a slow, uniform speed for approximately 20 minutes. Water flow through a net was measured by a current meter fastened in the center of the mouth.

Because of differences in the speed of the boat during hauling, in the length of time of hauling, and in the degree of clogging of the nets, the amount of water strained varied from haul to haul even for nets of the same size and construction. Also, nets of two different sizes were used during the surveys, the larger straining about two and a half times as much water as the smaller. For comparability, the numbers of eggs in all hauls were converted to the basis of numbers per standard volume of water strained per unit of depth fished. The standard chosen was 10 cubic meters of water strained per meter of depth fished, this particular value being used because it gave a factor of approximately 1 for the 1.0-meter net. In effect, this adjusts the number of eggs in a haul to an equivalent of the number under 10 square meters of sea surface. This area will hereafter be referred to as a "standard area." The procedures used and the resulting standard numbers for each station are listed in Ahlstrom (1948).

Our surveys did not begin early enough in either year to determine with certainty when pilchard spawning commenced off southern California. The first cruise (cruise 10) of the 1940 survey took place between March 5 and 14. Pilchard eggs were taken in every haul during this cruise, hence spawning was occurring throughout the area and must have commenced some time previously. Some indication of the length of the time involved may be inferred from the data on pilchard larvae. Specimens up to 16.5 mm. in length were collected during cruise 10, suggesting that spawning had been under way for at least a month. The first cruise of the 1941 survey took place two weeks later than in 1940 when the spawning season was even more advanced. The surveys, especially that of 1941, covered the end of the spawning season off southern California more adequately. During the last cruise of 1940, which took place May 27 to June 7, only about 1/20th as many eggs were taken as in midseason. Even fewer eggs were collected on the last two cruises of 1941, which took place after early June, their numbers being only about one-third of one per cent as many as were taken on the first cruise of this season.

In both seasons, pilchard spawning was widespread throughout the area surveyed off southern California, especially at the height of the

<sup>3</sup> Pilchard eggs do not ordinarily occur this deep: they are most concentrated in the upper 20 meters and very few are found below the 40 meter level.

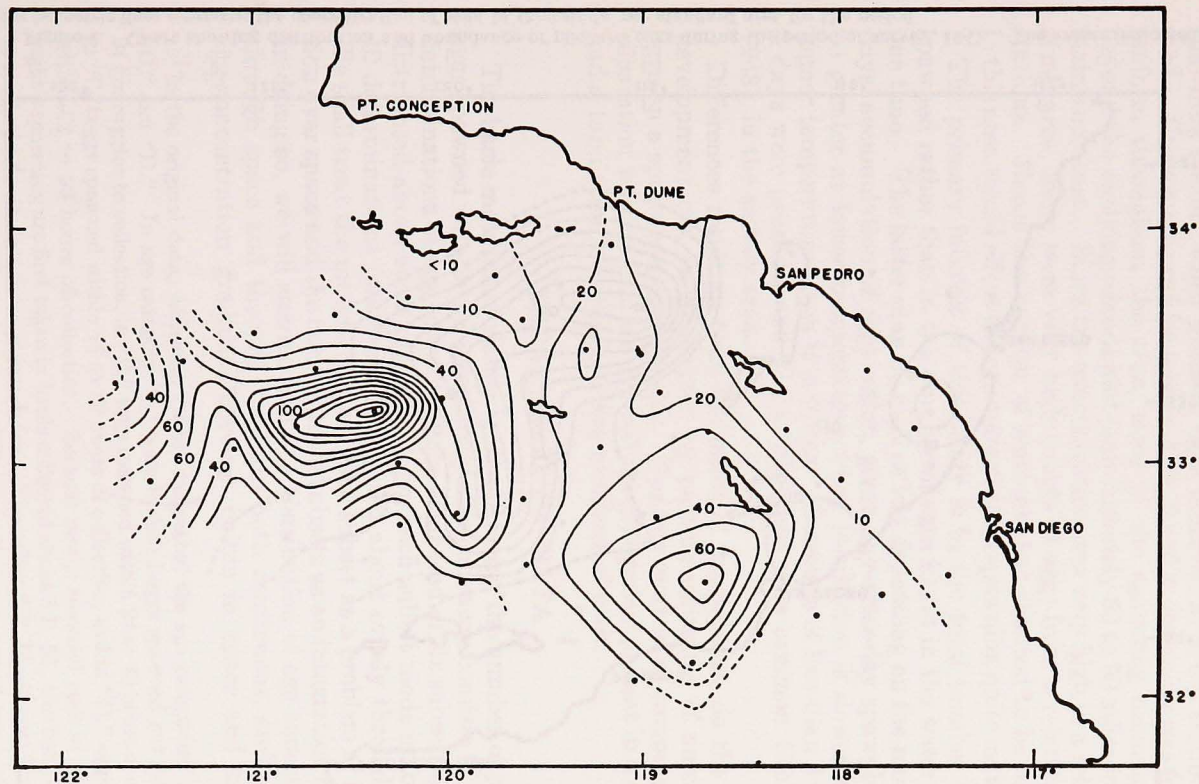


Figure 3. Chart showing distribution and abundance of pilchard eggs during the period of survey, 1940. The values indicated for isometric lines represent the concentration of eggs, in thousands, per standard area for the period.

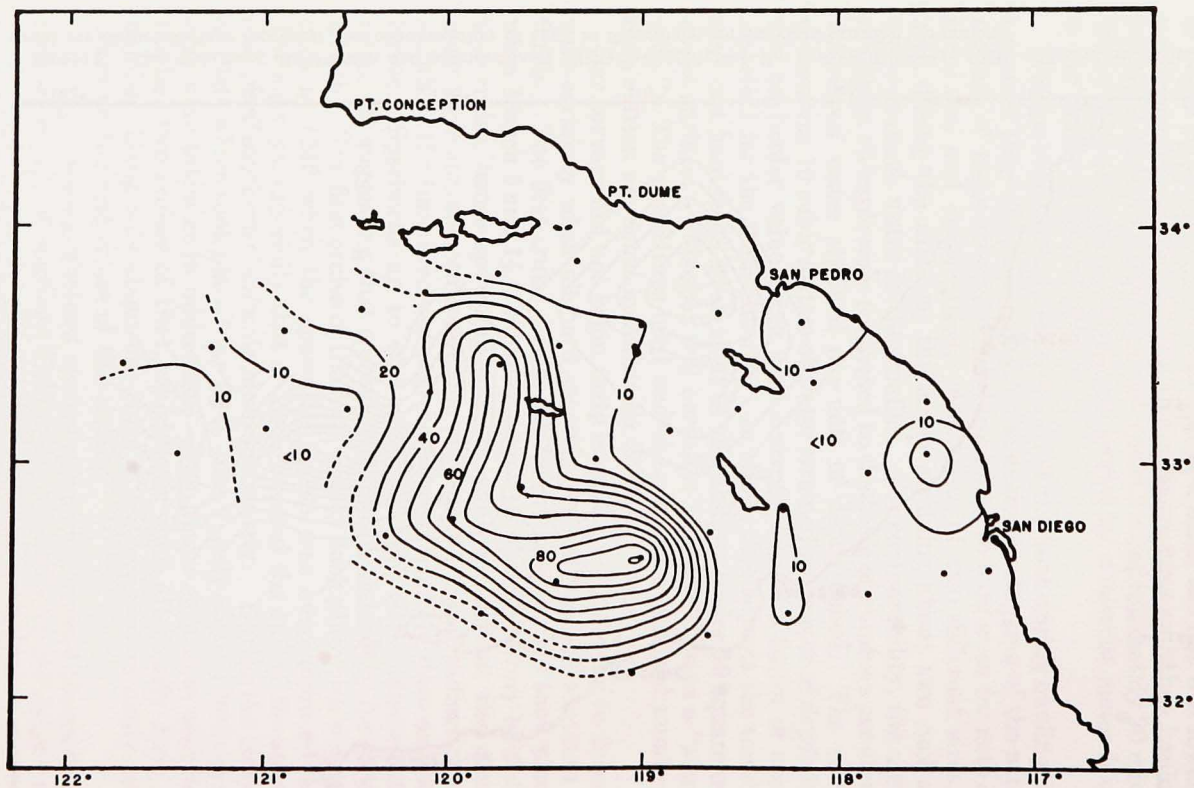


Figure 4. Chart showing distribution and abundance of pilchard eggs during the period of survey, 1941. The values indicated for isometric lines represent the concentration of eggs, in thousands, per standard area for the period.

spawning season, when pilchard eggs occurred at every station occupied. However, as can be seen in Figs. 3 and 4, spawning was far from uniform throughout the area, most of the spawning occurring just beyond the outlying islands and approximately 60 to 90 miles from the mainland coast. Since the concentration was very high in a relatively small area, there were very high counts of eggs from a relatively few stations. Hence the number of stations that chanced to be included in this area would effect the estimate of total spawning quite materially.

The primary interest of this study is in the total number of eggs spawned rather than in the number of eggs found in the water at any one time. The latter may be high or low depending on the number of days' accumulation of eggs which, given day-after-day spawning, will be greater at lower temperatures when incubation is slower than at higher temperature when it is faster. Extremes of less than 2 to over 4 days were possible, and  $2\frac{1}{2}$  to  $3\frac{1}{2}$  days were common (Ahlstrom, 1943), in the survey area.

Differences resulting from the effect of temperature on the rate of development can be eliminated by putting estimates of numbers of eggs on a spawned-per-day basis. For one method of computation, a separation was made of the several age-categories<sup>4</sup> present in a sample, and a tabulation was made of the numbers of each.

### TREATMENT OF DATA

The basic material available for determining the number of pilchard eggs spawned within the area surveyed each season consists of a set of determinations of egg concentrations composed of six subsets (cruises) distributed about evenly through time, each subset made up of about 40 determinations (stations) distributed about evenly through space. We shall treat the material within each subset as a problem of integration over space and the combining of subsets as an integration on time. In doing so, we will assume that the distribution of egg concentration through space and through time are both continuous, and that the egg-concentration gradients between points in space and time are

<sup>4</sup> In the original data, Ahlstrom (1948) designated the age categories "A," "B," "C," and "D." In age category "A" were included eggs spawned not longer than 24 hours prior to collection, in "B," eggs spawned within 24 to 48 hours of collection, in "C," eggs spawned within 48 to 72 hours of collection, and in "D," eggs spawned within 72 to 96 hours of collection. Because newly spawned eggs of the current night's spawning are first taken in hauls collected about 8 P. M., this hour was taken to mark the break between age categories. Thus after 8 P. M. all eggs of the previous night's spawning were classed as "B" eggs, all eggs spawned two nights before as "C" eggs, etc.



linear. Even with these simplifying assumptions there are several methods of estimating total numbers for a cruise or for a season.

*Definition of Symbols.* It will be convenient to adopt a series of symbols to represent various types of estimates and concepts.

$c_i$ : Any of several estimates pertaining to the number of pilchard eggs spawned per day per standard area of 10 square meters of sea surface, at the  $i$ -th station.

$c_i'$ : The total number of pilchard eggs in a standard haul or standard area at the  $i$ -th station.

$c_i''$ : The total number of pilchard eggs in a standard haul at the  $i$ -th station belonging to "complete age categories" (as the latter is defined elsewhere), *i. e.*,  $c'$  less incomplete age categories.

$c_{ia}$ : The number of eggs in the youngest complete age category at the  $i$ -th station.

$c_{ib}$ : The number of eggs in the next older complete age category at the  $i$ -th station.

$\bar{c}_i'$ : An estimate of  $c_i$ , derived by dividing  $c_i'$  by  $d_i$ .

$\bar{c}_i''$ : An estimate of  $c_i$ , derived by dividing  $c_i''$  by the number of age categories involved.

$c_{it}$ : The estimated number of eggs spawned in a standard area at the  $i$ -th station during the period of survey calculated by

$$c_{it} = D_i \frac{\Sigma c_i'}{h'd''}.$$

$C$ : The estimated number of eggs spawned within the survey area during the period of a cruise calculated by the formula

$$C = D \Sigma (w_i c_i).$$

$C'$ : The estimated number of eggs spawned within the survey area during the period of a cruise calculated by

$$C' = 649 \times 10^7 \frac{D}{d'h} \Sigma c_i' ;$$

$C_t$ : The estimated number of eggs in the survey area during the period of survey, *i. e.*, sum of  $C$  or  $C'$  for all cruises.

$C_T$ : The estimated number of eggs in the survey area during the spawning season.

$D$ : Period of cruise in days.

$D_t$ : Period of survey in days.

- $d_i$ : An estimate for each station of the time in days from spawning to hatching, determined from the relation between rate of development and temperature, the latter being the temperature at the 15-meter level at the station.
- $d'$ : An estimate for the period of a cruise of the average number of days from spawning to hatching, determined from the relation between rate of development and temperature, the latter being the average of temperatures at the 15-meter level for all stations of the cruise.
- $d''$ : An estimate for the spawning season of the average number of days from spawning to hatching based on the relation between rate of development and temperature, the latter being the average of temperatures at the 15-meter level for all of the stations included in the season estimate.
- $h$ : Number of stations occupied during a cruise.
- $h'$ : Number of times a station was occupied during the season.
- $w_i$ : Any of several weighting factors for the space, in standard areas, represented by the  $i$ -th station on a given cruise.
- $w_i'$ : The weighting factor that is proportional to the area of the polygon assumed to be represented by the  $i$ -th station.
- $w''$ : The weighting factor based on the assumption of equal weight for each station.
- $649 \times 10^8$ : Extent of survey area expressed as square meters of sea surface.
- $649 \times 10^7$ : Extent of survey area expressed in standard areas, *i. e.*, units of 10 square meters.

*Determination of  $c_i$ , the Number of Eggs Spawned per Day.* In estimating  $c_i$ , the number of pilchard eggs being spawned daily under a standard area (10 square meters) at the  $i$ -th station for a given cruise, two methods will be developed, one involving separation of age categories, the other not.

The latter method, being the simpler of the two, will be discussed first. The estimate is obtained by dividing  $c_i'$ , the total number of eggs in a standardized sample, by  $d_i$ , the estimated number of days from spawning to hatching, determined from the relation between rate of development and temperature. Estimates of  $c_i$  obtained by this method are designated  $\bar{c}_i'$ . The temperature of the 10-20-meter level was chosen because most pilchard eggs occur above 20 meters, with the largest concentration between 10 to 20 meters. We have used Ahlstrom's (1943) table 4 for values on the relation between temperature and the rate of development of pilchard eggs.

When basing estimates on individual age categories, one is faced with the problem of incomplete age categories resulting from collection being made while spawning or hatching was actively taking place. Since it can seldom be determined from an examination of the samples whether or not an age category is incomplete due to these causes, the decision has to be arrived at by other means.

From evidence based on the time of collection of newly-spawned eggs (incidence of stage I eggs, see Ahlstrom, 1943), most spawning seems to be confined to the period 8 P. M. to midnight. However, on comparing the number of "A" eggs with "B" eggs in the samples of the 1940 and 1941 seasons grouped by two-hour intervals, we find the following:

<i>Samples collected between:</i>	<i>Number of samples</i>	<i>Per cent that "A" age category is of the "B"</i>
8-10 P.M.	19	1
10-12 P.M.	23	3
0-2 A.M.	14	9
2-4 A.M.	23	34
4-6 A.M.	17	39
6-8 A.M.	28	125
8-10 A.M.	20	102

From this tabulation it is evident that the "A" age category was caught in markedly smaller numbers than the "B" category between 8 P. M. (when "A" eggs are first taken in hauls) and 2 A. M. During the next four hours "A" eggs, although somewhat more numerous, were still much less abundant than "B" eggs. The reason for this is not clear, since evidence, based on the stage of development reached by "A" eggs collected during this period, indicates that the majority of them must have been spawned several hours earlier. Although we cannot assign a definite reason for the under-representation of "A" eggs in samples collected during this period, it is too marked to be ignored. Consequently, "A" eggs collected between 8 P. M. and 6 A. M. were judged to be incompletely represented in the material, and these were excluded from estimates of  $c_i''$ ,  $\bar{c}_i''$ ,  $c_{ia}$ , and  $c_b$ .

When this criterion for incomplete categories is adopted, the classifications of eggs into A, B, C, and D as given by Ahlstrom (1943) must be revised. The former "A" category, collected between 8 P. M. and 6 A. M., is omitted; the "B," "C" and "D" categories taken in the same hauls become, for purposes of this report, "A," "B" and "C" categories, respectively. Those taken at other hours are unaffected.

An age category may also be incomplete due to a portion of the eggs having hatched. For distinguishing these, the assumption was made that if 80% or more of the eggs of an age category were in the stage immediately preceding hatching (stage XI; for a description see Ahlstrom, 1943), a portion of the eggs of the category would have hatched already; the category would be judged incomplete, and such groups of "C" or "D" age categories would be excluded from estimates of  $c_i''$ ,  $\bar{c}_i''$ ,  $c_{ia}$ , and  $c_{ib}$ .

Having eliminated incomplete age categories, an estimate may be based on "A" eggs alone, "B" eggs alone, or the mean of all complete age categories of eggs. These are designated as  $c_{ia}$ ,  $c_{ib}$ , and  $c_i''$ , respectively. No separate estimates were made for older age categories because, after elimination of incomplete age categories, there were only the two categories "A" and "B" left at many stations where the temperature was above 15° C, and less than three days were necessary for incubation.

*Determination of  $w_i$ , the Areas of Stations.* The concept for "area of station" is a convenience for one method of integration over space. It involves the idea that catch at a particular station is not only an estimate of egg concentration at that particular point but also an estimate of concentration over some area surrounding the particular point. Buchanan-Wollaston (1937) obviously was influenced by this idea in suggesting that a relative estimate of abundance of plankton in sets of hauls be made by weighting the catch at each station by a number proportional to the area of a circle whose radius was equal to the mean distance between the given station and adjacent ones. This method has two disadvantages: It is not always possible to define "adjacent station" satisfactorily, and the sum of the weighted catches is relative rather than absolute because the sum of the areas of the circles will not equal the area of the portion of the sea over which the stations were distributed.

However, if instead of circles a system of polygons be adopted, in which the polygon enclosing each station is erected by constructing the perpendicular bisectors of lines drawn from the station to each of all surrounding stations, it will be found that there is a unique polygon surrounding each station (except the peripheral ones). If the concentration at the station then be assumed to extend to the limits set off by each polygon, a system of weightings proportional to the areas of the polygons will result in a set of values which, when summed, will give results identical to those which would be obtained by linear interpolation between stations.

Such a system has the advantage that the sum of the weights equals the area surveyed; furthermore, it is completely objective, excepting possibly the treatment of peripheral stations, and as will be shown below, gives results essentially comparable to those from the conventional but less objective method of constructing isometric lines. Moreover, the polygons are easily constructed when one finds the perpendicular bisectors by swinging arcs of equal radii, the latter selected to be slightly greater than one half the distance between a given station and its neighbor. The areas of the polygons may be determined mathematically or by planimeter. With the latter, the work is about equal to that required when measuring the areas bounded by lines in an isometric system.

For these various reasons the above described system of polygons was adopted for our estimates involving area weightings, and in these the weight,  $w_i'$ , is the area of the polygon corresponding to the  $i$ -th station on a given cruise in 10-square-meter units, *i. e.*, "standard areas" as earlier defined. For peripheral stations, where coast lines were not involved, the outer boundary of the northern line was placed north from this line at half the average distance from this line to the next adjacent line to the south; the outer boundary of the southern line was placed south of the line at half the average distance from the line to the next adjacent line to the north; a station located at the end of a line was enclosed by assuming a station to be present on the line beyond the end station at a distance equal to the average distance separating stations on the line. The area of the resulting polygons was measured by a planimeter, using the coastline as a boundary where it overlapped a polygon.

Because one or more stations were omitted in some of the cruises, polygons were constructed for each cruise of the 1940 survey. An example of a complete one is given in Fig. 1. In 1941 no stations were omitted and a single system of polygons was erected for that season's station pattern, using the average position of occupying each station, as shown in Fig. 2.

It should be noted that the area surveyed during 1941 differed in several respects from the area surveyed the preceding season. The chief differences were: (1) Three stations (Nos. 26, 27, and 33) in the southwest portion of the 1940 survey pattern had no counterparts in the 1941 pattern, and (2) the southern line of four stations (Nos. 61 to 64) of the 1941 pattern had no counterpart in 1940. These stations without counterparts were omitted. The 1941 survey area then amounted to approximately 18,900 nautical square miles. The area surveyed during each cruise of the 1940 season was as follows:

Cruise Number	Area (nautical sq. mi.)	Factor to adjust area to 18,900 nautical sq. mi.
10	17,450	1.083
11	17,100	1.107
12	19,000	0.997
13	19,150	0.987
14	19,800	0.956
15	19,700	0.960

For comparison with each other and with the 1941 cruises, the 1940 cruises were adjusted to an area of 18,900 nautical square miles, using the above-listed factors. In so adjusting these cruises, we are assuming that the average abundance in a nonsurveyed portion was equal to the average abundances in a surveyed portion.

A more rapid, although less exact, estimate of egg numbers in the survey area may be made by weighting each station equally. Where this has been done, the symbol  $w''$  is used. For comparability of the results obtained by this method with others, the total was put on the basis of 18,900 nautical square miles ( $649 \times 10^8$  square meters) and stations without counterparts in both years were excluded. For a given cruise, the value of  $w''$  was determined by dividing  $649 \times 10^7$ , the number of standard areas of 10 square meters each in the survey area, by the number of stations occupied.

*Determination of D, the Time Period of a Cruise.* In integrating egg numbers over time, the same concepts were used as in integration over space; that is, the egg concentration observed at a particular time was assumed to represent an estimate for an adjacent period of time also. Specifically, each cruise was assumed to represent a certain period, the exact number of days in this period,  $D$ , being the sum of the number of days spent in making the cruise in question plus one half of the days intervening between the end of the preceding cruise and the start of the cruise under consideration, plus one half the time interval in days between the end of the cruise under consideration and the start of the following cruise. The time interval preceding the first cruise of a season was assumed to represent half as long a time as the interval between the first and second cruises. Similarly, the time interval following the last cruise of a season was assumed to represent half as long a time as the time interval between it and the preceding cruise.

By this method of determining  $D$ , the result is a computed total which is equivalent to one employing a linear interpolation in time, just as was true with respect to space for the determinations of  $w$ .

*Estimates of Total Eggs Spawned.* The foregoing sections give four methods of determining  $c$ ; and two for determining  $w$ ; hence, there are

eight possible estimates available for each cruise. By adding the appropriate estimates for the six cruises for each season, estimates of  $C_t$ , the number of pilchard eggs in the survey area during the period of the survey have been obtained; these are given in Table I. However, because spawning had commenced in the survey area some time prior to the start of the first cruise of each season, the estimates of pilchard eggs in the area during the period of the survey did not represent the number of eggs spawned during the complete spawning season in either year. From the number and sizes of larvae taken during the first cruise of each season, it was estimated, by a procedure to be reported upon in a subsequent paper, that 11% in 1940 and 17% in 1941 of the spawning had taken place prior to the period covered by the survey estimate. The season totals after adding these percentages are denoted by  $C_T$ .

*Simplified Procedure for Estimating the Amount of Spawning in the Survey Area Based on Average Conditions During a Cruise.* The methods previously described require the separate determination of  $c_i$  and  $w_i'$  for each station. A computation that does not require this is:

$$C' = 649 \times 10^7 \frac{D}{d'h} \Sigma c_i', \quad \text{where use is made of}$$

the sum of the standard haul totals for all stations occupied during the cruise, put on a per-day basis by dividing this value by the number of days required for the eggs to develop from spawning to hatching at the average temperature of the 15-meter level during the cruise, and  $C_i$  by summing the  $C'$  for all cruises of the season.

*Determination using Isometric Lines.* In practice, when using the conventional method based on isometric lines, the number of lines must be kept relatively few, and there is some subjectivity in drawing the lines, hence the pattern follows linear interpolation approximately rather than precisely. Nonetheless, for showing distribution, the isometric chart is nearly indispensable, and it can also serve for estimating total quantities. Having a single chart showing the abundance during each year, our estimation was done for the season as a whole rather than by individual cruises. The values for each station, given in Figs. 3 and 4, represent estimates of the number of eggs spawned during the period of survey in a standard area, *i. e.*,  $c_i$  as defined previously. The area between lines was determined by planimeter. The estimates obtained by this method and given in the last column of Table I are similar to those obtained by the other methods.

TABLE I. COMPARISON OF ESTIMATES OF NUMBER OF PILCHARD EGGS, IN BILLIONS, SPAWNED IN AREA SURVEYED OFF SOUTHERN CALIFORNIA DURING SPAWNING SEASONS OF 1940 AND 1941

Cruise No. and Date	Estimates Based on Formulation $C = D\Sigma(w_i c_i)$ Using								$C'$	Average	Using Isometric Lines
	$w_i' c_{ia}$	$w_i' c_{ib}$	$w'' c_{ia}'$	$w'' c_{ib}$	$w_i' c_i''$	$w'' c_i''$	$w_i' c_i'$	$w'' c_i'$			
1940:											
10 (III:5-14)	29869	28224	33082	29590	27459	30220	25848	27626	27843	28862	
11 (III:20-28)	22540	32566	23074	26588	31144	28614	30765	28431	29073	28088	
12 (IV:4-14)	22231	15722	25256	17182	23457	24704	22567	23367	23311	21977	
13 (IV:22-V:3)	51830	54046	62716	61900	52452	62007	48299	58145	57933	56592	
14 (V:10-21)	29611	30314	30826	35290	31779	34377	30260	33009	35257	32303	
15 (V:27-VI:7)	3126	2321	3674	2745	2832	3240	2410	2702	2717	2863	
Survey total, $C_t$	159207	163193	178628	173295	169123	183162	160149	173280	176134	170685	180929
Season total, $C_T$	178884	183363	200706	194713	190026	205800	179943	194697	197903	191781	203291
1941:											
17 (III:21-30)	97185	61531	95282	62432	78515	77368	68024	66983	69009	75148	
19 (IV:14-25)	14932	67001	13542	61549	39649	36452	38509	35979	38350	38440	
21 (V:8-17)	10584	16642	11464	17245	12801	13385	13625	14206	15230	13909	
23 (VI:2-12)	5510	4053	5442	3716	4601	4425	4669	4474	4582	4608	
25 (VI:25-VII:3)	442	103	409	110	268	256	254	244	242	259	
27 (VII:21-28)	33	5	36	10	28	22	23	28	21	23	
Survey total, $C_t$	128686	149335	126175	145062	135862	131908	125104	121914	127434	132387	128335
Season total, $C_T$	155043	179922	152018	174773	163689	158925	150728	146884	153535	159502	154620



*Comparison of Estimates.* The estimates of the number of pilchard eggs spawned in the survey area during the spawning seasons of 1940 and 1941 are tabulated in Table I. The lowest estimate for the 1940 season is 87% of the highest and 93% of the average estimate; the lowest estimate for the 1941 season is 82% of the highest and 92% of the average estimate.

For each season, a group of four estimates differs from a second group only in the method of determining  $w$ . The estimates based on  $w'$  attempt to weight each station according to the portion of the survey area represented by it; those based on  $w''$  give equal weight to the stations. A comparison of these estimates, in billions of eggs, follows:

Value of $c$ used	1940 Season			1941 Season		
	Estimate using			Estimate using		
	$w'$	$w''$	$w'/w''$	$w'$	$w''$	$w'/w''$
$c_a$	178,884	200,706	.891	155,043	152,018	1.020
$c_b$	183,363	194,713	.942	179,922	174,773	1.029
$\bar{c}''$	190,026	205,800	.924	163,689	158,925	1.030
$\bar{c}'$	179,943	194,697	.924	150,728	146,884	1.026
Average	183,054	198,979	.920	162,345	158,150	1.027

The two groups of estimates agree more closely for the 1941 season than for the 1940 season. This is not surprising in view of the fact that the 1940 station pattern was more irregular than the 1941 pattern.

A comparison of estimates, in billions of eggs, using different methods of determining  $c$ , shows the following:

1940					
	$c_a$	$c_b$	$\bar{c}''$	$\bar{c}'$	average
$w'$	178,884	183,363	190,026	179,943	183,054
$w''$	200,706	194,713	205,800	194,697	198,979
Average	189,795	189,038	197,913	187,320	191,016
1941					
	$c_a$	$c_b$	$\bar{c}''$	$\bar{c}'$	average
$w'$	155,043	179,922	163,689	150,728	162,345
$w''$	152,018	174,773	158,925	146,884	158,150
Average	153,530	177,347	161,307	148,806	160,248

All estimates of  $c$  agree quite well for the 1940 season, less well for the 1941 season. The poorer agreement for the latter season may be partly the result of fewer samples having been taken during the period of more intensive spawning during this season. The  $\bar{c}'$  estimates are

somewhat lower than those based on age categories: 3% for 1940, 9% for 1941.

The simplified procedure,  $C'$ , which can be calculated much more rapidly than any other given in the paper, is within 4% of the average estimate for each season, and would alone suffice as a determination of the number of eggs spawned in the area during these seasons.

### RELIABILITY OF THE ESTIMATES

A number of counts and measurements enter into the process of computing the total number of eggs spawned in the survey area. Each of the counts or measurements represents more or less imperfectly the true value of the measurement sought. In other words, each contributes to the variability or statistical error of the final total. The two principal mensuration categories are 1) the concentration of eggs at points in time and space, and 2) the locations of the points in time and space. In addition to these there are the approximations involved in interpolations.

*Variability of Egg-concentration Data.* Sampling of the survey area was done at points (stations) that were about evenly distributed rather than randomly distributed through space and time. Likewise, the egg concentrations that were being sampled were not randomly distributed. The nonrandomness of the distribution in space may be seen by referring to Figs. 3 and 4, and in time by comparing the cruise totals given in Table I. The concentration of eggs at a given point was measured by making a plankton haul, sorting and counting subsamples of the catch, and subsequently converting the count to represent the number of eggs in a standard area.

Due to the nonrandom nature of the distribution of measurement points and of the eggs themselves, the statistical error of individual hauls cannot be measured by root mean sums of squares of deviations from the grand mean. Instead, each datum on egg concentration must be considered as a sample drawn from a different universe, or alternatively, the entire collection of egg concentration might be regarded as samples stratified in time and space. However, the latter alternative is not a promising approach to the study of reliability unless values both of the egg concentrations and of their variances can be expressed as functions of time and space (or in relation to some other frame of reference such as possibly physical, chemical or biological qualities of the sea water in which the eggs were taken and which might determine their concentration through their influence on the act of spawning). Accordingly we shall examine the data from the standpoint suggested in the first alternative, that is, as if each con-

centration datum had been drawn from its own peculiar population.

Fortunately, the possibility of the need for this study was foreseen in planning the surveys, when "paired hauls" were included in the design of observation for five of the 1940 cruises. By the method of analysis of variance of egg counts in a group of the paired hauls of 1940, selected for containing more than a certain minimum number of eggs, Silliman (1946) already has studied the variances ascribable to differences between stations (time-place of hauls), between members of a paired haul at the same station and between samples counted from the same haul, as well as "haul-order," "sample-order" and interactions. Several features of his findings are of interest here.

First, the variability (mean square) between stations was about twelve times as high as the variability between hauls, proving statistically the point discussed above, that hauls taken at different stations cannot be considered as samples drawn from the same population.

Second, due to the markedly skewed distribution of numbers of eggs per sample, Silliman found it necessary to use the logarithms of the egg counts, which he found were nearly normally distributed. In this use of logarithms, Silliman followed Winsor and Walford (1936), and Winsor and Clarke (1940). The latter, in considering the counts of the various plankton organisms in each of twelve oblique hauls made in succession and considered as a sample of the same body of water, had found that the range of variation was very different for the different groups of plankton organisms, the range being roughly proportional to the mean for each group. They also observed a tendency for the numbers of different kinds of organisms in any haul to be high or low together, indicating factors common to the whole haul affecting all groups in the same direction. They concluded that "under these circumstances, it is both convenient and theoretically advantageous to use the logarithm of the catches rather than the catches themselves" (Winsor and Clarke, 1940: 4).

Since Silliman carried through his analysis in logarithms, his estimate, say of 0.024 for variance between hauls, is the variance of the logarithms of egg-count, not of the counts themselves. Similarly, the square root of the variance, 0.155, is the standard deviation of the logarithms of egg-counts. The antilog of this standard deviation, 1.429, stands in relation to the egg-counts themselves as a ratio by which the egg-count is multiplied or divided to get the upper and lower one-standard-deviation limits, respectively, of the sampling variability of the egg-count.

These upper and lower limits are not equidistant from the actual value of the egg-count. If we let the logarithm of egg-count equal zero, its antilog will equal one, and the upper limit is .429 while the

lower is  $1 - 1/1.429 = .300$  from the value (one) of the egg-count. This asymmetry, of course, is due to the fact that when the distribution of the logarithms of a variable is normal, the distribution of the variable itself is skewed, the curve falling more sharply to the left and more gradually to the right of the mode. Attending this asymmetry, of course, is the difference between the antilog of the mean of the distribution of the logarithms of the variable and the mean of the distribution of the variable itself. In this case the antilog of the mean of the logarithms of the egg-counts is one, as stated. The mean of the actual egg-counts themselves is approximately 1.06. This will be recognized as a specific example of the familiar fact that the geometric mean (antilog of the mean of logarithms of numbers) is always somewhat lower than the arithmetic mean of a series of numbers. It is of interest here, because it suggests that if the sampling variation of numbers of eggs in hauls is such that their logarithms are distributed normally, the mean of these numbers will be somewhat higher than its most probable or modal value.

Of greater significance is the fact that the sampling variation of numbers of eggs in hauls is proportional to the number in the haul. In the case of one egg in the haul, we have seen that the upper and lower one-standard-deviation limits are .429 and .300, respectively, from the value one. For 100 eggs the corresponding limits are 42.9 and 30.0, respectively. The one-standard-deviation range is exactly one hundred times as great.

Our estimates of the number of eggs spawned into the survey area are essentially sums of numbers, each of which has the sampling variability of the type discussed above. As indicated by the example given, the sampling variability is large, and as expressed by the standard deviation is proportional to the number itself. Given this kind and degree of variability in the numbers summed, what is the variability of the sum itself?

Silliman (1946: 82), in giving a table of fiduciary limits for means of egg-counts, evidently had this question in mind. But this method of computing fiduciary limits, using logarithms, gives a result that is valid only for geometric means, or for arithmetic means in the special case where the number in the egg-counts is identical for all items entering the mean. It would be a useful approximation for expressing the sampling variation of the mean of a number of hauls made at nearly the same time in nearly the same locality where it might be expected that the number of eggs per haul would be nearly uniform among the several hauls.

With the present case, where the numbers entering the sums range through several orders of magnitude, and with the sampling varia-

bility proportional to the number itself, it is obvious that the larger numbers contribute vastly more to the variability of the sums than the smaller numbers, and this must be reflected in the sampling variation of the sums of such numbers.

It is proposed, therefore, to employ the fact that the variance of a sum of independent variates is equal to the sum of the variances of the variates, and to estimate the variance of a variate from the variance of the logarithms of that variate. In our case the variate is the number of eggs per haul at a station, and we will use the values for the variance of the logarithm of the number of eggs per haul computed by Silliman (1946) for a group of 24 paired hauls of the 1940 survey. Since this group is not necessarily a random sample of the hauls comprising our material, the use of Silliman's results gives only an approximation of the true variances. To be discussed later are the several respects in which the nonrandom nature of the 24 hauls selected by Silliman may influence our estimate.

In estimating the variance of the number of eggs in a haul, it is convenient to take advantage of the fact that it is the standard deviation and not the variance (or standard deviation squared) that is proportional to the number of eggs in the haul. Since the sampling distribution of the logarithms of the numbers is normally distributed, the root-mean-square of the deviation from the mean of the sampling distribution of the corresponding numbers themselves will be an estimate of their standard deviation. This distribution is skewed, of course, and the individual standard deviations may not be given the same interpretation as when computed for a normal distribution. However, when a number of sets of samples are drawn from a population, the distribution of the means, and therefore the sums of the sets, approaches normality even though the population from which drawn be markedly skewed. It will be assumed that this holds true for our sums of egg-counts, where the count itself is analogous to the individual in a sample and where the sampling distribution of hauls is analogous to the population from which the sample is drawn; our problem, then, is to find the standard deviation of the distribution of  $x$  where the distribution of  $\log x$ , denoted by  $Lx$ , is normal with the estimated standard deviation of  $s_{Lx}$ . A general mathematical solution would involve extensive analysis, but, having to deal with only several values of  $s_{Lx}$ , it is fairly simple to erect a curve of distribution of  $x$  which corresponds to the curve of distribution of  $Lx$  for a given  $s_{Lx}$  through use of tabular values for the normal frequency function such as the one given in Hodgman (1944: 200-204). Given this distribution of  $x$ , its estimated standard deviation,  $s_x$ , may be readily computed by the root-mean-square method.

This was done for haul counts, the mathematical expectation of the logarithm of which is zero (one egg per haul) and values of  $s_{Lx}$  appropriate (according to Silliman, 1946) to cases where (1) the egg number in our data was estimated from one haul from which one sample was counted ( $s_{Lx} = 0.176$ ) and (2) where it was estimated from two hauls from which two samples were counted ( $s_{Lx} = 0.1183$ ). Also, it was done for two arbitrary values:  $s_{Lx} = 0.10$  and  $s_{Lx} = 0.25$ . The results are as follows:

$s_{Lx}$	Antilog. of $s_{Lx}$	$s_x$	$\bar{x}$	$s_x/\bar{x}$
0.1760	1.500	0.4505	1.0856	0.415
0.1183	1.313	0.274	1.038	0.264
0.1000	1.259	0.224	1.027	0.218
0.2500	1.778	0.732	1.180	0.620

The arbitrary values of  $s_{Lx}$  were included to explore the possibility of a linear relationship between  $s_{Lx}$  and  $s_x$  or associated values. The relation between antilog  $s_{Lx}$  and  $s_x$  seems to approach linearity and is approximately defined by

$$s_x = .944 (\text{antilog } s_{Lx}) - .965. \quad (1)$$

This formula, calculated from the co-ordinates antilog  $s_{Lx} = 1.5$ ,  $s_x = 0.4505$  and antilog  $s_{Lx} = 1.313$ ,  $s_x = 0.274$ , gives the value  $s_x = .227$  for  $s_{Lx} = .1000$  and  $s_x = .729$  for  $s_{Lx} = .250$  as compared with the above-listed values for  $s_x$  obtained directly by the root-mean-square calculation. The agreement is as close as would be expected from consideration of computational accuracy, and linear interpolation for other values of  $s_x$  may be confidently used within the range  $s_{Lx} = 0.1$  and  $s_{Lx} = 0.25$ .

In estimating the total numbers of eggs spawned into the survey area, the counts of eggs were multiplied by certain factors before getting their sums. Since the standard deviation of the number caught in a haul,  $s_x$ , is proportional to  $\bar{x}$ , the number caught in a haul, as earlier explained,

$$s_x = s_x \cdot \bar{x},$$

where the mathematical expectation of  $x$  equals one, then

$$s_{ax} = a \cdot \bar{x}s_x,$$

where  $a$  represents any weighting factor.

Furthermore, where such weightings have been applied to summed sets of egg-counts, such as weighting the egg numbers for a cruise according to the time-period represented, the standard deviation of

the sum,  $s_s$ , is determined from the standard deviations of the individual haul counts,  $s_{x_i}$ , by the equation:

$$s_s = (s_{x_1}^2 + s_{x_2}^2 + \dots + s_{x_n}^2)^{\frac{1}{2}},$$

and, if the counts be multiplied by weighting factors  $w_i$ ,

$$s_{w_s} = (w_1^2 s_{x_1}^2 + w_2^2 s_{x_2}^2 + \dots + w_n s_{x_n}^2)^{\frac{1}{2}};$$

also, where the  $w_i$  are all equal,

$$s_{w_s} = w(s_{x_1}^2 + s_{x_2}^2 + \dots + s_{x_n}^2)^{\frac{1}{2}}.$$

Inasmuch as all of the estimated totals of numbers of eggs spawned in the survey area are based on the same original hauls, the portion of their variability attributable to sampling variations of hauls must be similar, unless the weightings employed were markedly different. It is probably sufficient, therefore, to estimate the haul variability associated with the estimate of  $C_t$  in the column headed  $w_i'c_i'$  in Table I.

We shall let the several values for  $w_i'c_i'$  equal  $x_1, x_2$ , etc. There were three types of haul-sample combinations in the 1940 and 1941 survey data for which the following values were used:

Station value based on:	Series Designation	$V_{Lx}$ (from Silliman)	$s_{Lx}$	Antilog. $s_{Lx}$	$s_s$ (from Eq. 1)
1 haul, 2 samples	$x'$	.0275	.1658	1.465	0.418
1 haul, total count	$x''$	.024	.1549	1.429	0.384
2 hauls, 4 samples	$x'''$	.014	.1183	1.313	0.274

The computation for the estimate of total eggs spawned for one season's survey,  $C_t$ , was computed by the formula

$$C_t = D_1 \Sigma x_i + D_2 \Sigma x_i + \dots + D_n \Sigma x_i,$$

where  $D$  is a weight proportional to the number of days' spawnings represented by a cruise and its subscripts the first, second, etc., cruises of the season.

The computation of the standard deviation for sampling variation for the sum of eggs of each cruise,  $s_c$ , then is:

$$s_c = (s_{x_1}^2 + s_{x_2}^2 + \dots + s_{x_n}^2 + s_{x_1}''^2 + s_{x_2}''^2 + \dots + s_{x_n}''^2 + s_{x_1}'''^2 + s_{x_2}'''^2 + \dots + s_{x_n}'''^2)^{\frac{1}{2}};$$

and the computation for  $s_{C_i}$ , the standard deviation for sampling variation for the sum of the cruises for the season then is:

$$s_{C_i} = (D_1^2 s_{C_1}^2 + D_2^2 s_{C_2}^2 + \dots + D_n^2 s_{C_n}^2)^{\frac{1}{2}},$$

where the subscripts of  $D$  and  $C$  denote the first, second, etc., cruises of the season.

The resulting values for  $C_i$  and  $s_{C_i}$  are given in Table II in two parts: one part consisting of the contributions from stations whose  $w'\bar{c}'$  is more than 20 and the other part whose  $w'\bar{c}'$  is less than 20, designated in the table as high-concentration stations and low-concentration stations, respectively. The high-concentration stations have egg-counts lying within the range studied by Silliman and the low-concentration stations have counts lying below this range.

It was mentioned earlier that Silliman's estimate of variance,  $V_{Lx}$ , was on a nonrandom sample of our hauls, and the most important departure from randomness was his elimination of samples with low egg-counts. In the latter, both the sampling variability of hauls and of samples drawn from the catches made by the hauls for the counting of eggs must be larger relative to the count itself than for those with large counts. Therefore, the contribution to the standard deviation of the sum for all stations is underestimated for the small hauls. However, for both 1940 and 1941 this contribution is so small that even a 100% change in the standard deviation contributed by the low-concentration stations would change the standard deviation of the survey area sum by considerably less than 1%.

There are other respects in which the statistics for  $V$ , as given by Silliman, may not estimate correctly the variability of our data. Undoubtedly  $s_x$  varies directly with  $x$  because there are a number of factors operating on a haul and on the measurements pertaining to the haul as a whole. For our particular hauls there are several such factors that can be recognized readily. For instance, Silliman found "haul-order" and "sample-order" variances that could be traced to faults in the mechanics of handling the material caught by the hauls. The discrepancies were such as to tend to be cancelled upon pooling the two members of paired hauls in 1940; the mechanical faults themselves were rectified in 1941 procedures. Hence, these do not materially affect the  $s_x$  values employed herein. They serve as examples of factors, unrelated to the actual concentrations of eggs in sea, which can affect the number of eggs supposed to represent the concentration in the sea at a certain time and place.

Similar mechanical faults may affect the hauling process itself. In our hauls the most likely fault is the variation in the amount of water



strained at the levels of maximum concentration of eggs. It will be recalled that the hauls were oblique and that the level at which the net was towing at any particular time was controlled by the length of towline and the angle of stray, as reported by Ahlstrom (1948). The length of towline could be controlled precisely enough, but the angle of stray was a function of speed of the ship relative to the water mass acting upon the net and could be controlled only very approximately. Operational experience has shown that sometimes the net towed more than twice as much at one level than at another through the series of some 20 levels that made up each oblique haul. Records of vertical distribution of eggs (Ahlstrom, 1948) also proved that eggs were sharply stratified vertically, with by far the highest concentrations occurring within the upper 20 meters. However, the townet was drawn from about 70 meters to the surface and only a part of the tow took place in the upper 20 meters. While the total amount of water strained during the haul was registered by current meter and computed from initial and final readings, this could reveal nothing as to the amount of water strained in the upper egg-rich layers. This no doubt explains Silliman's (1946: 81-82) failure to find any reduction of variability between paired hauls after both were "standardized" by taking into account the measurements of water passing through the net in each haul.<sup>5</sup> This, we believe, is the principal reason that the sampling variability in our hauls is large and is proportional to the number of eggs in the haul instead of being in accordance with the Poisson series, as might be expected *a priori* and according to which its standard deviation would be proportional to the square root of the number.

Due to improvement in 1941, both in equipment and in procedure, we believe that the sampling variation attributable to the making of hauls was substantially reduced in 1941 as compared with 1940, and in using the values for  $V$  based on 1940 material we tend toward placing a maximal value on the standard deviation for the 1941 estimation.

Also, the sampling variation attributable to the drawing of the samples for counting the eggs was lowered in 1941 as compared with 1940, both by using a better fractioning device and by drawing a larger sample. Accordingly, the standard deviation is overestimated as concerns this factor also.

It is our opinion that the three principal respects in which our

<sup>5</sup> The type of haul employed in 1940 and 1941 was designed to reach the maximum depth at which larvae as well as eggs are found. For taking only eggs to reduce sampling variability of hauls, it would have been better to haul the nets from about 20 meters to the surface. However, this would have caused a systematic under-representation for eggs and even more serious faults with respect to larvae.

TABLE II. VARIABILITY OF ESTIMATE OF TOTAL EGGS SPAWNED IN SURVEY AREA

<i>Category</i>	<i>No. of Sta. (h)</i>	<i>Est. of eggs spawned (C<sub>i</sub>)</i>	<i>Stand. dev. of est. (s<sub>C<sub>i</sub></sub>)</i>	<i>Coefficient of variation (100s<sub>C<sub>i</sub></sub>/C<sub>i</sub>)</i>
1940:				
High-concentration stations	83	150,534	7,539	5.01
Low-concentration stations	89	9,619	467	4.85
All stations	172*	160,153	7,553	4.71
1941:				
High-concentration stations	45	117,437	12,056	10.27
Low-concentration stations	73	7,656	557	7.28
All stations	118*	125,093	12,069	9.64

\* Includes all stations at which pilchard eggs were taken, excludes stations with 0 concentration of eggs.

material is not typical of the material from which our value for  $V$  was taken have the net effect of somewhat underestimating the standard deviation of the estimate for 1940 and overestimating it for 1941.

One may conclude, therefore, that a true value for the standard deviation for 1940 would be somewhat over 7,553, and for 1941 somewhat under 12,069, perhaps not over 10,000 billion in either year. The difference of the survey area estimates of the two years is approximately 35,000 billion eggs, and hence the estimates for the two seasons are somewhat over two standard deviations apart. Thus, as far as sampling variability of hauls is concerned, the observed difference between the two seasons is demonstratively significant.

*Dimensions in Time and Space.* The dimensions in time and space which entered into the estimates as weighting factors were measured with sufficient accuracy, in a physical sense, to be negligible in relation to other sources of variability. However, time enters into the computation fundamentally through its use to measure the duration of the incubation period of the eggs, in the estimates based on  $c'$ , which, it will be recalled, is the catch of eggs divided by the time in days from spawning to hatching and which varies with temperature. The accuracy of such estimates depends, then, on the reliability of the determination of the relation between temperature and development and on the faithfulness with which the temperatures observed at the collecting point reflect the temperature during the entire period of development up to the collecting time. There is opportunity for

systematic error in both of these mensuration elements. For this particular set of estimates, any such systematic error would have very little effect on the result, because the scale of development rate was determined from the very same set of data (including the measurement of temperature) as was used for our estimates. In effect, this adjusts the mensuration instrument for observational error, and the relative numbers of eggs estimated to have been spawned should not be affected except for the variability associated with predicting the incubation time from the regression line of time to reach stage XI on temperature at the 10 to 20 meter stratum, as given by Ahlstrom (1943: table 5). Since the standard error of such predictions, given in terms of the reciprocal of the logarithm of the temperatures, is equivalent to approximately 7% of the incubation period, and since some hundreds of such estimates of  $\bar{c}'$  are involved in the total survey estimate, this error is negligible. It may be noted in passing, however, that the use of Ahlstrom's (1943) rate-of-development schedule in calculating other sets of data may easily bring in systematic error if there be different relations between the temperature and the vertical distribution of pilchard eggs.

Unlike the estimates employing  $\bar{c}'$ , those using  $c''$ ,  $\bar{c}''$ ,  $c_a$  and  $c_b$  are entirely free from error connected with rate of development. This would be true not only for the present set of observations but for any set where the eggs are separable into daily age categories.

In general, therefore, the estimates of total eggs spawned in the survey area are negligibly affected by errors in the factors used for space and time weightings.

*Errors Introduced by Interpolations.* The estimates in this report assume continuity and linearity with space and time in the distribution of egg concentration. Both of these assumptions would tend toward validity as the spacing of stations in space and time becomes infinitesimal. The question of errors involved in these assumptions, therefore, becomes a question of whether or not the actually rather wide intervals in space and time completely destroy the validity of the assumptions or only introduce variability.

Two approaches are possible. In one, we may examine the effects of changes in spacing. In the other, we may examine evidence inherent in the patterns described by our sets of data.

Entering the first approach, it is, of course, impossible to experiment with closer spacing without collecting more material. Actually, a special set of material was collected in 1941, consisting of catches of pilchard eggs along a line of stations spaced  $1\frac{1}{2}$  to  $3\frac{1}{2}$  miles apart; hence, about 5-10 times more closely spaced than in the cruises here

reported upon. Studies of the special set have not yet proceeded far enough to be included here in full. Gross inspection of the data, however, bears out the assumption of continuity. It suggests also that the assumption of linearity in the egg-concentration gradients between stations about 20 miles apart (as in the regular set) instead of the  $1\frac{1}{2}$  to  $3\frac{1}{2}$  miles apart (as in the special set) is materially but not disastrously faulty. The short-distance changes in the concentration gradient appear to be such as to be described better by curves than by straight lines. If the numbers of eggs per haul be plotted as ordinates on a base line scaled in miles, points connecting the tops of ordinates at about 20-mile intervals enclose a polygon whose area is not very dissimilar to that of the polygon constructed by connecting the tops of all ordinates. The material and its study are not sufficient to deduce any general relation between the spacing of stations and the egg numbers computed therefrom.

Although it is not possible, beyond this special set of data, to experiment with closer spacing of stations, we have experimented with wider spacing. For this purpose, the stations were divided into subsets by two procedures: (1) using alternate stations and (2) using alternate

TABLE III. COMPARISON OF ESTIMATES OF THE NUMBER OF PILCHARD EGGS, IN BILLIONS, AS DETERMINED FROM ALTERNATE STATIONS, FROM ALTERNATE LINES OF STATIONS, AND FROM THE TOTAL STATIONS—ALL ESTIMATES MAKING USE OF THE FORMULATION  $C = D\Sigma (w'' c_i'')$ .

Sta. No.	Total Sta.	Alt. Sta. I	Alt. Sta. II	Sta. Lines	
				1-3-5	2-4 & Adj. Sta.
1940: 10	30220	22835	33576	19894	41836
11	28614	33989	23574	26750	31565
12	24704	26386	23023	16037	34609
13	62007	63236	60840	23300	107542
14	34377	32513	37954	30933	40453
15	3240	2996	3470	3676	2659
Survey Total, $C_i$	183162	181955	182437	120590	258664
Season Total, $C_T$	205800	204444	204985	135494	290634
1941: 17	77368	86960	67552	46239	114320
19	36452	22955	50700	23355	51862
21	13385	13269	13481	19734	5887
23	4425	4830	3999	3638	5352
25	256	418	108	298	197
27	22	18	46	15	49
Survey Total, $C_i$	131908	128450	135886	93279	177667
Season Total, $C_T$	158925	154759	163718	112384	214057

lines of stations, *i. e.*, lines 1, 3, and 5 in one group, lines 2 and 4 in the other. The estimate for each group was based on the average of the several age categories in each sample,  $\bar{c}''$ , and the stations were given equal weight in the integration over the area,  $w''$ . The results are given in Table III.

The two estimates for each season, based on groups of alternate stations, are nearly identical for the 1940 season and are within 5% of each other for the 1941 season. On the other hand, the estimates based on alternate lines of stations are quite dissimilar; the sum based on lines 1, 3, and 5 is only about half as large as that based on lines 2 and 4 in both 1940 and 1941. The high sum for the latter resulted from both lines passing through relatively high concentrations, whereas the low sum for the former resulted from only one of the three lines passing through the area of high concentration. Obviously, the spacing of lines much farther apart than the approximately 30-mile intervals of 1940 or 1941 would make a system of estimates based on interpolations relatively inaccurate.

Such empirical demonstrations pertain only to a particular set of data treated in a particular way and have little predictive value for results from other sets or from other possible treatments.

The other approach, that of examining evidence inherent in our set of data, may be of more general significance. First, as to continuity, inspections of charts for individual cruises, in which the distribution of eggs is expressed by drawing lines that are isometric with respect to egg concentration, show the irregularity of contours to be only slightly greater than the ones given in Figs. 3 and 4, which are for the means of the six cruises in 1940 and 1941, respectively. In all of them, the concentration changes in a more or less regular and gradual manner. There is always a zone (sometimes two) of high concentration, usually including several stations, with the concentration falling off in all, or nearly all, directions away from this zone.

There is not only evidence of this continuity but also some indication of the shape that a frequency surface would have, were one to erect at the station locations a set of ordinates proportional in height to the egg concentrations. Such a frequency surface would have relatively gentle contours in the zone of high concentration, with the contours falling off more sharply away from this zone and then becoming more gentle again when approaching the quite low concentrations at some distance from the zone of maximum and near much of the periphery of the survey area in our charts. Such a surface resembles a normal frequency surface in being convex upward near the mode, inflecting at some distance from the mode to become concave upward at a still greater distance. In fact, one may test the distribution of hauls, with

respect to egg-counts, against the distribution expected from a normal frequency surface according to a method suggested by Sette (1943: 195-198). For the 1940 survey, such a test gives a probability of 0.57 that the observed hauls could have been drawn from a normal frequency surface or from a set of several such surfaces. Hence, an integration of egg numbers for our surface on the assumption of linear gradients would be analogous to representing a univariate frequency distribution by a frequency polygon, and our method of treating the station concentration as uniform in the "surface-area" (*i. e.*, uniform from the station to a point midway to adjacent stations) is analogous to a univariate frequency histogram. Since the areas enclosed by the frequency polygon and by the frequency histogram are identical when both are drawn through identical points, the frequency histogram will differ from the frequency curve by the same amount as does the frequency polygon. The area of the latter, when class intervals are relatively small, will not differ much from that of the frequency curve itself. However, for large class intervals, where, by chance, no class center is near the mode, a serious underestimate would result.

Translated to an appraisal of the situation in our set of data, this means that our estimates are in error, through the assumption of linearity, if the stations were so far apart that none chanced to lie at or very near the point of maximum concentration. This would produce a systematic error tending toward lowering the estimate, first, because the maximum values contribute most to the sum, and second, because the normal frequency surface, which our set of data apparently simulates, has a much smaller area of above-average than below-average surface. Since it is unlikely that every cruise would have a station at or very near the point of maximum concentration, our estimates probably lie below the true values in both years. Furthermore, with stations spaced approximately the same distance apart in both seasons, the underestimation probably would be greater in the season in which the eggs were the less widely distributed. From Figs. 3 and 4, it would seem that the 1941 total would be the one that is more severely underestimated. We know of no presently available means of appraising the degree of underestimation and suggest that a study of the spacing of stations in relation to the distribution of eggs be given a high priority in future work in this field.

Comparison of charts of individual cruises, with isometric lines drawn with respect to egg concentration, with similar charts of temperature, salinity, oxygen content, phosphate content and dynamic surface, suggest that the zones of egg concentrations are located in a relatively stable relation to combinations of some aspects of these water qualities. It is suggested that efforts to relate the physical and

chemical system to the distribution of eggs would not only be illuminating for the factors influencing spawning, but also would provide important material for designing cruise plans to improve quantitative estimates of spawning.

This discussion thus far has related exclusively to interpolations in space. Practically the same considerations and the same uncertainties exist with respect to time. Spawning may be expected to rise from low rates at the beginning of the season to a peak later in the season and then fall away to low values toward the end of the season. This seems true of the two seasons of the present set of data. With only six points to describe the course of rise and fall in each, there is a hazard of failure to chance upon near-peak values, similar to the hazard of failing to occupy stations in the area of near-maximal concentration, with the result that effects on the resulting estimates would be similar.

*General Appraisal of Reliability of Estimates.* It has been seen that sampling variability of hauls introduces an uncertainty of estimate of about 5% at the one-standard-deviation limit; measurements of time and space have introduced only negligible error; interpolations, due to the wide intervals between stations in space and cruises in time, have probably introduced a systematic error, perhaps greater in one season than the other, but in both tending toward lowering the estimate.

In addition to this, there is another error systematically operating in the same direction, which so far has not been considered. That is, the effect of mortality during the incubation period. Our estimates were based on eggs in several age-categories. All groups, except eggs just spawned, have been subjected to some mortality, and therefore the estimates tend not to be of the number of eggs spawned but of the number of eggs part way through embryonic development.

The independent estimates of  $c_a$  and  $c_b$  throw some light on the probable amount of mortality and hence of the effect of mortality on our estimates. Since eggs of age-category "B" are one day older, on the average, than those of age-category "A," the former will have been subject to one day's more mortality than the latter. Therefore, one would expect the estimate of  $c_b$  to be lower than that of  $c_a$ . On the contrary, our estimate of  $c_b$  is larger than that for  $c_a$  in 1941, and the two are about equal in 1940. We can find no explanation for this, other than that the error of estimate is sufficiently great to obscure the effect of mortality. If so, the error due to the latter must be less than that due to the former and is probably negligible. Whether this probable negligibility extends to age-categories "C" and "D" is difficult to determine, because there are fewer data (due to incomplete-

ness of daily age-categories), and random variability of the sum has larger play.

Finally, it should be noted that neither season's estimate embraced the whole spawning season in the survey area, and the survey itself included only a portion of the spawning range. Estimates of the percentage loss of eggs from our totals, due to incomplete coverage of the season within the survey area, have been given, but their reliability is undetermined. How far the estimate of totals for the survey area falls short of the total of spawning throughout the entire spawning range cannot even be estimated. Indeed, it is not determinable from our data whether the survey area included comparable parts of the total spawning activity in the two years. Hence, no conclusions may be reached as to the absolute or even relative amount of spawning in the two years, 1940 and 1941.

### CONCLUSIONS

The study has brought to light some aspects of the problem of estimating spawning productivity, from quantitative plankton tows, that may be of significance to further advances in this field.

1. The estimation of total eggs in an area is possible by several methods, differing in detail, but when based on the same fundamental assumptions, the more easily computed ones give results similar to the more laborious ones for the particular station patterns employed in 1940 and 1941.

2. The station patterns were more nearly adequate in the spacing between stations on the station line than in the spacing between station lines, the latter not being susceptible to substantial increase without materially altering the estimates.

3. Although the sampling variability of our plankton hauls is large for individual hauls, it becomes moderate for the sum of such hauls.

4. Since the error connected with station spacing may be substantial and is probably systematic, the opportunity for advances in the understanding of factors influencing the distribution of the eggs is probably greater than that afforded by further refinements intended to decrease haul variability.

### REFERENCES

AHLSTROM, E. H.

1943. Studies on the Pacific pilchard or sardine (*Sardinops caerulea*). 4.—Influence of temperature on the rate of development of pilchard eggs in nature. Spec. Sci. Rep. U. S. Fish Wildl. Serv., 23. 26 pp. (Mimeo.).
1948. A record of pilchard eggs and larvae collected during surveys made in 1939 to 1941. Spec. Sci. Rep. U. S. Fish Wildl. Serv., 54. 76 pp. (Mimeo.).



BUCHANAN-WOLLASTON, H. J.

1937. Two Technical Notes. (2) Estimating the total or the mean number of organisms under a given area of sea or a given area of sea bottom. *J. Cons. int. Explor. Mer*, 12 (3): 334.

CLARK, F. N.

1939. Measures of abundance of the sardine, *Sardinops caerulea*, in California waters. *Fish Bull.*, Sacramento, 53. 45 pp.

HODGMAN, C. D.

1944. Hand book of chemistry and physics. 28th ed. Chemical Rubber Publishing Co., Cleveland. 2571 pp.

SCOFIELD, E. C.

1934. Early life history of the California sardine (*Sardina caerulea*), with special reference to distribution of eggs and larvae. *Fish Bull.*, Sacramento, 41. 48 pp.

SETTE, O. E.

1943. Biology of the Atlantic mackerel (*Scomber scombrus*) of North America. Part 1. Early life history. *Fish Bull.*, U. S. Fish Wildl. Serv., 38: 149-237.

SILLIMAN, R. P.

1946. A study of variability in plankton townet catches of Pacific pilchard (*Sardinops caerulea*) eggs. *J. Mar. Res.*, 6 (1): 74-83.

SVERDRUP, H. U. AND STAFF

1944. Oceanographic observations on the "E. W. Scripps" cruises of 1940. *Rec. Obser. Scripps Inst. Oceanogr.*, 1 (3): 161-248.

1947. Oceanographic observations on the "E. W. Scripps" cruises of 1941. *Rec. Obser. Scripps Inst. Oceanogr.*, 1 (4): 249-407.

WINSOR, C. P. AND G. L. CLARKE

1940. A statistical study of variation in the catch of plankton nets. *J. Mar. Res.*, 3 (1): 1-34.

WINSOR, C. P. AND L. A. WALFORD

1936. Sampling variations in the use of plankton nets. *J. Cons. int. Explor. Mer*, 11 (2): 190-204.