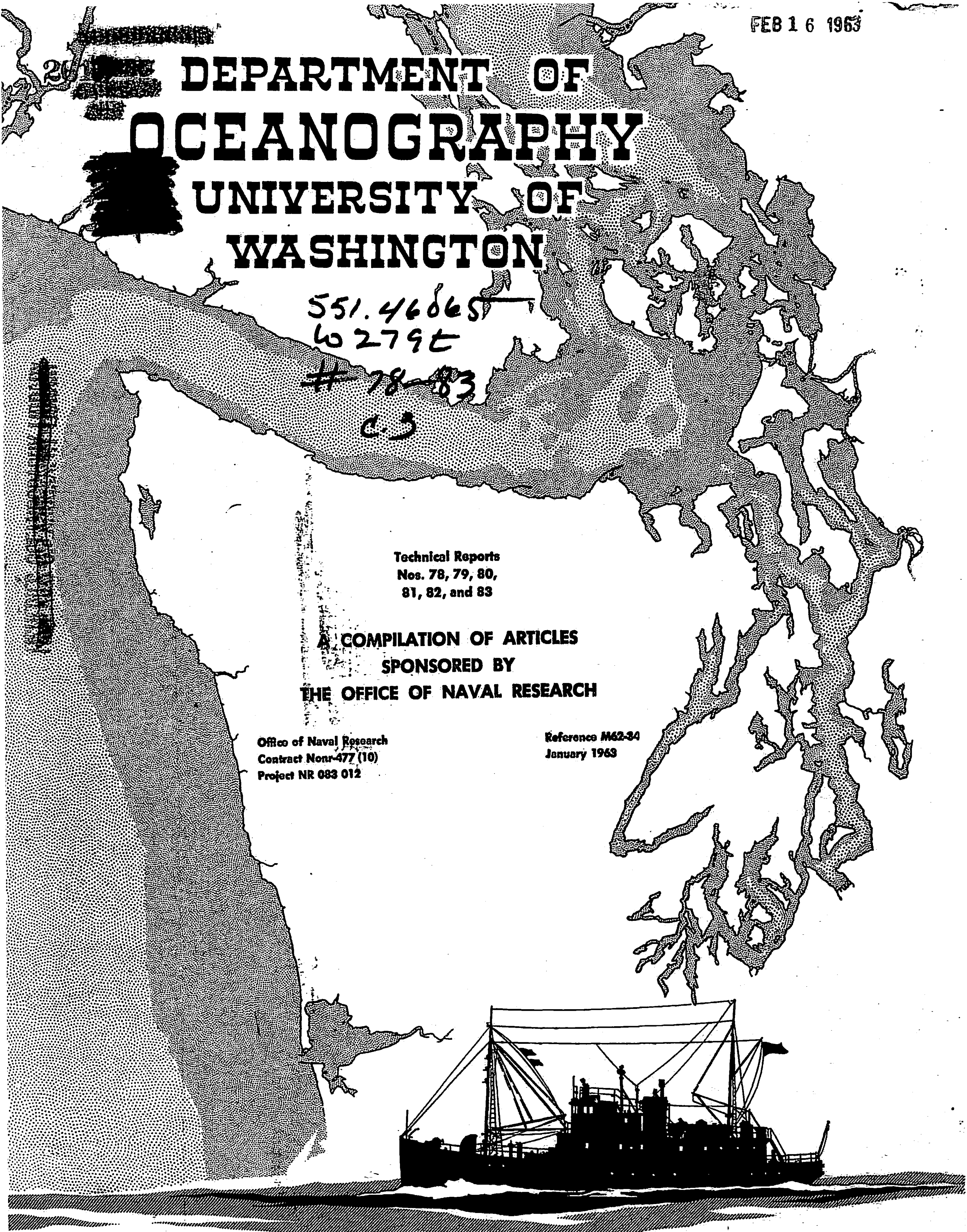

DEPARTMENT OF OCEANOGRAPHY UNIVERSITY OF WASHINGTON

Technical Reports
Nos. 78, 79, 80,
81, 82, and 83

A COMPILATION OF ARTICLES SPONSORED BY THE OFFICE OF NAVAL RESEARCH

Office of Naval Research
Contract Nonr-477 (10)
Project NR 083 012

Reference M62-34
January 1963

SEATTLE 5, WASHINGTON

UNIVERSITY OF WASHINGTON
DEPARTMENT OF OCEANOGRAPHY
Seattle 5, Washington

Articles Sponsored by the Office of Naval Research:

Technical Report No. 78

A SIMILARITY SOLUTION FOR CIRCULATION IN AN ESTUARY, by Maurice Rattray, Jr. and Donald V. Hansen. Journal of Marine Research, 20(2):121-133.

Technical Report No. 79

INTERPOLATION ERRORS AND OCEANOGRAPHIC SAMPLING, by Maurice Rattray, Jr. Deep-Sea Research, 9:25-37.

Technical Report No. 80

DISTRIBUTION OF PHYSICAL PROPERTIES BELOW THE LEVEL OF SEASONAL INFLUENCE IN THE EASTERN NORTH PACIFIC OCEAN, by Maurice Rattray, Jr., Cuthbert M. Love, and Diane E. Heggarty. Journal of Geophysical Research, 67(3):1099-1107.

Technical Report No. 81

RESEARCH ACTIVITIES AT THE UNIVERSITY OF WASHINGTON, DEPARTMENT OF OCEANOGRAPHY, by Karl Banse, Joe S. Creager, Richard H. Fleming and Clifford A. Barnes, Maurice Rattray, Jr., Francis A. Richards. The First National Coastal and Shallow Water Research Conference, pp. 724-736.

Technical Report No. 82

NET ZOOPLANKTON AND TOTAL ZOOPLANKTON, by Karl Banse. Rapp. et Proc.-Verb., 153(36):211-215.

Technical Report No. 83

A SIMPLE SEMIAUTOMATIC REAGENT DISPENSER, by Ralph W. Riley and Francis A. Richards.

Office of Naval Research
Contracts Nonr-477(10)
Project Nr 083 012

Reference M62-34
January 1963

RICHARD H. FLEMING
Chairman

Deep-Sea Research, 1962, Vol. 9, pp. 25 to 37. Pergamon Press Ltd. Printed in Great Britain

Interpolation errors and oceanographic sampling*

MAURICE RATTRAY, JR.

(*Received* 26 *August* 1961)

Abstract—Formulae are derived which can be used to give an interpolated value for an oceanographic variable and to estimate the accuracy of the interpolation. The interpolated value is obtained by taking the average of the values obtained from two Lagrange interpolation polynomials : one utilizing two points above and one below the depth of interpolation; and the other, one above and two points below this depth. The error of an interpolated value is composed of two parts, one part which is due to measurement error of the observations, and another part which is the error of the interpolation itself. The difference in the values given by the two interpolation polynomials provides a measure of the error of interpolation. Expressions for the errors resulting from measurement errors, given as the ratio of the standard deviation of the interpolated values to the standard deviation of the individual measurements, depend only on the sampling and interpolation depths. Utilization of these formulae permits the determination of optimum sampling programmes which, with a minimum number of observations, yield the distribution of a variable to a prescribed accuracy.

INTRODUCTION

IN RECENT years, especially with the advent of machine computation of routine oceanographic data, a number of interpolation schemes have been evolved for obtaining values of the oceanographic variables at standard depths (FOFONOFF and TABATA, 1958; KILMER and DUXBURY, 1959; FROESE, 1960). There has been a question as to which scheme is the most suitable for this purpose and also as to whether any is equal to the classical graphical methods of interpolation. Meanwhile, there is growing recognition that error in the actual measured values of the oceanographic variables must be considered in any interpretation of oceanographic data (STRICKLAND, 1958; WOOSTER and TAFT, 1958). It seems appropriate at this time to consider the complete problem of obtaining desired data within a specified degree of accuracy. To achieve this goal it is necessary to evaluate the effects of both interpolation and measurement errors and then, based on these results, to develop the requirements for a satisfactory sampling programme.

Factors affecting the accuracy of measurement of the oceanographic variables will not be considered here. Further, the accuracy of any measurement will be assumed to be known and fixed. Thus, for example, a standard deviation can be attached to each reported salinity or temperature observation. While a change in accuracy of measurement will affect the sampling and interpolation programmes, it will not affect the criteria for evaluation of errors presented in this paper.

Although the interpolation scheme, described herein, seems to give a better interpolated value over a wider range of conditions than others that have been proposed, its main justification for selection is that it enables an error of interpolation to be determined.

*Contribution Number 242 from the Department of Oceanography, University of Washington.

INTERPOLATION

Oceanographic data are frequently analysed by means of interpolation procedures yielding errors which may have far-reaching consequences in interpretations of the data. It therefore seems essential to make valid estimates of these errors, particularly since the use of large computing machines makes it practicable to do so.

The method used for determining the interpolated value and its error is based on Lagrange's parabolic interpolation polynomial (BUCKINGHAM, 1957) for three positions of the observed values. Let us take values of the dependent variable y_1, y_2, y_3, y_4 to be known at successive positions x_1, x_2, x_3, x_4 where $x_4 > x_3 > x_0 > x_2 > x_1$. Then the value of y_0 at a position x_0 lying between x_2 and x_3 is

$$y_0 = p_{ijk}(x_0) + R_{ijk}(x_0) \tag{1}$$

with i, j, k equal to either 1, 2, 3, or 2, 3, 4, respectively, and where, from Lagrange's interpolation formula,

$$p_{ijk}(x) = \frac{(x - x_j)(x - x_k)}{(x_i - x_j)(x_i - x_k)} y_i + \frac{(x - x_i)(x - x_k)}{(x_j - x_i)(x_j - x_k)} y_j + \frac{(x - x_i)(x - x_j)}{(x_k - x_i)(x_k - x_j)} y_k, \tag{2}$$

and the errors are given by :

$$R_{ijk}(x) = \frac{(x - x_i)(x - x_j)(x - x_k)}{3!} y'''(\xi_j), \qquad x_i \leqslant \xi_j \leqslant x_k. \tag{3}$$

In practice, with either choice for i, j, k, the polynomial can be calculated but the error term is unknown.

Two estimates of the interpolated value at position x_0 may be determined by means of equation (1) : the first, utilizing two points above and one below x_0; the second, one point above and two below. Use of the resulting formulae will permit making a decrease in the error as well as an estimate of its magnitude. This is accomplished by taking a linear combination of the two obtained expressions, namely :

$$y_0 = [r\, p_{123}(x_0) + (1 - r)\, p_{234}(x_0)] + [r\, R_{123}(x_0) + (1 - r)\, R_{234}(x_0)] \tag{4}$$

where the error for a value of y_0 calculated by this formula is zero when

$$r(x_0) = r_0(x_0) \equiv 1 \Big/ \left[1 + \frac{(x_0 - x_1)\, y'''(\xi_2)}{(x_4 - x_0)\, y'''(\xi_3)}\right] \tag{5}$$

It is not possible to determine r_0 for each interpolation depth. Instead, it is necessary to use the same value of $r = \bar{r}$ for all depths and to accept some error in the interpolation. It remains to determine an appropriate value for $\bar{r}$ and then to obtain expressions which can be used to estimate the interpolation error. Before selecting a value for $\bar{r}$ it is useful to consider the possible values for r_0 and their effect on the corresponding interpolation errors.

In the majority of cases $y'''(\xi_2)$ and $y'''(\xi_3)$ will have the same sign and consequently r_0 will lie in the range $0 < r_0 < 1$. In the exceptional circumstance that $y'''(\xi_2)$ and $y'''(\xi_3)$ have different signs, there must be a zero of y''' in the intervening range and (except for highly irregular distributions) both $y'''(\xi_2)$ and $y'''(\xi_3)$ will be small. Thus, the error terms in the separate interpolation formulae will also be small. In these cases the difference in values of the two interpolation polynomials will be less than those for adjoining depths and the relative weighting resulting from the selection of $\bar{r}$ will have a lesser effect on the interpolation error. If the distribution is highly irregular between sampling points, no method of interpolation can give good results. However, when the criteria developed in this paper are applied, these irregular distributions will be made evident by large interpolation errors which indicate that the sample spacing should be decreased. Therefore, the significant range of r_0 for purposes of error determination is $0 < r_0 < 1$ and a choice of $\bar{r} = 0{\cdot}5$ will be useful for the largest number of conditions. In the remaining sections of this paper, this value will be used exclusively although this selection is not necessary for the validity of the arguments.

ERROR OF INTERPOLATION

The error of interpolation, $R(x_0)$, is obtained from the error term in equation (4) by substituting $\bar{r}$ for r :

$$R(x_0) = \bar{r}\, R_{123}(x_0) + (1 - \bar{r})\, R_{234}(x_0) \tag{6}$$

and with the help of equations (1)–(3) the error becomes

$$R(x_0) = \{p_{123}(x_0) - p_{234}(x_0)\}\left\{\frac{\bar{r}\,(x_0 - x_1)\,y'''(\xi_2) - (1 - \bar{r})\,(x_4 - x_0)\,y'''(\xi_3)}{(x_0 - x_1)\,y'''(\xi_2) + (x_4 - x_0)\,y'''(\xi_3)}\right\} \tag{7}$$

$$= \{p_{123}(x_0) - p_{234}(x_0)\}\{\bar{r} - r_0(x_0)\}. \tag{8}$$

The error is thus equal to the difference in values obtained by the two Lagrange interpolation formulae times a factor depending upon the third derivatives. In the second factor of equation (7), $\bar{r}$, $(1 - \bar{r})$, $(x_0 - x_1)$, and $(x_4 - x_0)$ are positive. If $y'''(\xi_2)$ and $y'''(\xi_3)$ have the same signs, then the absolute value of this factor must always be less than unity. On the other hand, if there is part of the range in which these derivatives have different signs, it will mean that the third derivative has a zero in this range, and (unless the curve is behaving wildly) both the third derivative and the interpolation error will generally be smaller than they are at neighbouring points. Such anomalous cases will be evident where the difference, $p_{123}(x_0) - p_{234}(x_0)$, is markely less for one position than for adjacent ones. It follows that in a series of interpolations the interpolation error will usually be bounded by the difference in the values of the interpolation polynomials. Only occasionally will the error be larger than this amount, and in these cases it will still be bounded by the difference in interpolation polynomials for the adjacent locations, provided that the interpolation interval is maintained.

Since $\{p_{123}(x_0) - p_{234}(x_0)\}$ is determined for each interpolation, it is apparent from equation (8) that a statistical measure of the interpolation error depends on the statistical properties of $(\bar{r} - r_0)$. With a proper choice of $\bar{r}$, the average value of this factor will be small, and with $\bar{r} = 0{\cdot}5$ it will rarely be more than $0{\cdot}5$. To make more definite statements about the error it would be necessary to evaluate the third derivatives, which is not feasible in practice. Thus, in order to arrive at some estimates of the statistical behaviour of $(\bar{r} - r_0)$, a synthetic example is constructed as follows :

Consider a depth interpolation problem with sampling and interpolated depths as shown in TABLE 1. With respect to this table, two assumptions are necessary :

(1) Wherever a sampling and an interpolated depth are listed as identical, the sampling depth is assumed, for purposes of error analysis, to be somewhat less than the listed value. This assumption seems justified because of the effect of wire angle. (2) It is assumed that the ratio of third derivatives can be satisfactorily estimated by considering the vertical distribution of properties to vary linearly with the logarithm of the depth, as suggested by TULLY (1957), and by taking ξ_2 and ξ_3 equal to x_2 and x_3, respectively. Then

$$y'''(\xi_2)/y'''(\xi_3) \approx x_3^3/x_2^3$$

as used in the table. On the basis of the foregoing assumptions, values of r_0 are tabulated for each interpolation depth. It is not important that these values be individually representative, but it is expected that the statistical distribution of the r_0's obtained in this example will be reasonably typical of what occurs in nature.

Table 1. Analysis of errors for a typical scheme of sampling and interpolation depths

Sampling Depths				Interp'n Depth	Error Analysis Data					
x_1	x_2	x_3	x_4	x_0	$\frac{x_0 - x_1}{x_4 - x_0}$	$\frac{x_3^3}{x_2^3}$	r_0	σ_p^2/σ^2	σ_Δ^2/σ^2	Σ/σ^2
0	10	20	30	10	0·50	8·0	0·20	1·0	0·0	0·0
10	20	30	50	20	0·33	3·4	0·48	1·0	0·0	0·0
20	30	50	75	30	0·22	4·6	0·50	1·0	0·0	0·0
30	50	75	100	50	0·40	3·4	0·42	1·0	0·0	0·0
50	75	100	150	75	0·33	2·4	0·56	1·0	0·0	0·0
75	100	150	200	100	0·25	3·4	0·56	1·0	0·0	0·0
100	150	200	250	150	0·50	2·4	0·45	1·0	0·0	0·0
150	200	250	300	200	0·50	2·0	0·50	1·0	0·0	0·0
200	250	300	500	250	0·20	1·7	0·75	1·0	0·0	0·0
250	300	500	750	300	0·11	4·6	0·67	1·0	0·0	0·0
250	300	500	750	400	0·43	4·6	0·33	1·3	2·0	− 0·4
300	500	750	1000	500	0·40	3·4	0·42	1·0	0·0	0·0
300	500	750	1000	600	0·75	3·4	0·28	0·7	0·4	− 0·1
300	500	750	1000	700	1·33	3·4	0·18	0·8	0·2	+ 0·1
500	750	1000	1500	800	0·43	2·4	0·50	0·8	0·1	0·0
750	1000	1500	2000	1000	0·25	3·4	0·56	1·0	0·0	0·0
750	1000	1500	2000	1200	0·56	3·4	0·35	0·8	0·6	− 0·1
1000	1500	2000	—	1500	—	—	—	—	—	—
1500	2000	—	—	2000	—	—	—	—	—	—
					Average =		0·45			
					Take $\bar{r}$ =		0·5			

In principle, on the basis of these assumptions, it would be possible to arrange the depths such that values of r_0 for the different depth ranges would be nearly the same. Practically, however, this could not be done with actual station data, and these synthetic results can only be used to indicate the expected range of r_0 and to show conditions which favour either small or large values of r_0. For computation of interpolated values at standard depths, based on these tabulated sampling depths, $\bar{r} = 0{\cdot}5$ is approximately the average value of r_0 and the standard deviation, σ_r, of $(\bar{r} - r_0)$ is estimated to be 0·17. For values of $(\bar{r} - r_0)$ normally distributed about zero, for 95 per cent confidence limits, the interpolation error would lie between

$$\pm\, 0{\cdot}33\,\{p_{123}(x_0) - p_{234}(x_0)\}. \tag{9}$$

Only the factor 0·33 is determined from this synthetic example, and from considerations of the limits of the interpolation error its value would appear reasonable for actual station data. It will be used in the following sections. A different choice for this quantity does not affect the further argument, and in practice the interpolation error should be kept sufficiently small, so that the actual value of this factor is unimportant.

ERROR OF MEASUREMENT

The expression for the estimated error of an interpolated value of a variable given above was based on exact, measured values of the variable. Actually, the measured values also contain errors which will be reflected as additional errors in the interpolated values. In order to show separately the effects of the measurement and of the interpolation errors on the interpolated value of a variable, the measurement errors are assumed to be normally distributed, independent of one another and of $\{\bar{r} - r_0(x_0)\}$. If σ represents the standard deviation of the individual measurements, and if σ_0, σ_p, σ_R, σ_Δ, and σ_r represent, respectively, the standard deviations of y_0, $\{\bar{r}\, p_{123}(x_0) + (1 - \bar{r})\, p_{234}(x_0)\}$, $R(x_0)$, $\{p_{123}(x_0) - p_{234}(x_0)\}$ and $\{\bar{r} - r_0(x_0)\}$, then in consequence of equations (4), (6) and (8) the standard deviation, σ_0, is given by

$$\sigma_0^2 = \sigma_p^2 + \sigma_R^2 + \Sigma \tag{10}$$

where

$$\sigma_R^2 = \{p_{123}(x_0) - p_{234}(x_0)\}^2 \sigma_r^2 + \{\bar{r} - r_0(x_0)\}^2 \sigma_\Delta^2 \tag{11}$$

and

$$\begin{aligned}
\frac{\Sigma}{\sigma^2} = {} & \\
& 2\,[\bar{r} - r_0(x_0)] \Bigg[(x_0 - x_2)^2 (x_0 - x_3)^2 \left\{\frac{(1 - \bar{r})}{(x_4 - x_2)^2 (x_4 - x_3)^2} - \frac{\bar{r}}{(x_1 - x_2)^2 (x_1 - x_3)^2}\right\} \\
& + \frac{(x_0 - x_3)^2}{(x_2 - x_3)^2} \left\{\frac{(1 - \bar{r})(x_0 - x_4)^2}{(x_2 - x_4)^2} - \frac{\bar{r}\,(x_0 - x_1)^2}{(x_2 - x_1)^2}\right\} \\
& + \frac{(x_0 - x_2)^2}{(x_3 - x_2)^2} \left\{\frac{(1 - \bar{r})(x_0 - x_4)^2}{(x_3 - x_4)^2} - \frac{\bar{r}\,(x_0 - x_1)^2}{(x_3 - x_1)^2}\right\} \\
& + \frac{(2\bar{r} - 1)(x_0 - x_1)(x_0 - x_4)}{(x_3 - x_2)^2} \left\{\frac{(x_0 - x_3)^2}{(x_2 - x_1)(x_2 - x_4)} + \frac{(x_0 - x_2)^2}{(x_3 - x_1)(x_3 - x_4)}\right\}\Bigg].
\end{aligned} \tag{12}$$

The term Σ expresses the correction which must be applied since

$$\{\bar{r}\, p_{123}(x_0) + (1 - \bar{r})\, p_{234}(x_0)\} \quad \text{and} \quad \{p_{123}(x_0) - p_{234}(x_0)\}$$

are not strictly independent, but both depend upon the measured values of the variable. The ratio Σ/σ^2 may be either positive or negative and, with reasonably selected sampling and interpolation locations, will be small. Typical values for Σ/σ^2 are shown in TABLES 1 and 4. Since this term is small, it will be neglected in the following discussions. Then the expression for the standard deviation of the interpolated value of a variable is approximated by

$$\sigma_0^2 = \sigma_p^2 + \{\bar{r} - r_0(x_0)\}^2 \sigma_\Delta^2 + \{p_{123}(x_0) - p_{234}(x_0)\}^2 \sigma_r^2. \tag{13}$$

The last term has already been considered in the previous section and represents the direct contribution of the interpolation error. The first two terms on the right side of the equation express the effects of the measurement errors. Except for unusual circumstances, $\sigma_\Delta{}^2$ will be less than $\sigma_p{}^2$ and when multiplied by $\{\bar{r} - r_0(x_0)\}^2$ the contribution of the second term will be negligible with respect to the first. As an example, typical values for $\sigma_p{}^2/\sigma^2$ and $\sigma_\Delta{}^2/\sigma^2$ are shown in TABLES 1 and 4.

Thus, the desired approximate equation is

$$\sigma_0{}^2 \cong \sigma_p{}^2 + \{p_{123}(x_0) - p_{234}(x_0)\}^2 \sigma_r{}^2, \tag{14}$$

and the contribution of the measurement errors to the square of the standard deviation of the interpolated value of the variable is given by the square of the standard deviation of the interpolation polynomial. In terms of the standard deviation due to measurement error, σ, this contribution is :

$$\frac{\sigma_p{}^2}{\sigma^2} = \bar{r}^2 \frac{(x_0 - x_2)^2 (x_0 - x_3)^2}{(x_1 - x_2)^2 (x_1 - x_3)^2} + \left\{\bar{r} \frac{(x_0 - x_1)(x_0 - x_3)}{(x_2 - x_1)(x_2 - x_3)} + (1 - \bar{r}) \frac{(x_0 - x_3)(x_0 - x_4)}{(x_2 - x_3)(x_2 - x_4)}\right\}^2$$
$$+ \left\{\bar{r} \frac{(x_0 - x_1)(x_0 - x_2)}{(x_3 - x_1)(x_3 - x_2)} + (1 - \bar{r}) \frac{(x_0 - x_2)(x_0 - x_4)}{(x_3 - x_2)(x_3 - x_4)}\right\}^2 + (1 - \bar{r})^2 \frac{(x_0 - x_2)^2 (x_0 - x_3)^2}{(x_4 - x_2)^2 (x_4 - x_3)^2}. \tag{15}$$

The ratio of these standard deviations is therefore completely determined by the sampling and interpolation locations and the value of $\bar{r}$.

It is evident both from equation (15) and the results given in TABLE 1 that the ratio $\sigma_p{}^2/\sigma^2$ can be less than unity. In this circumstance, the error in an interpolated value of a variable can be made less than that of an individual measurement provided that the difference in the two interpolation polynomials is made small compared to the standard deviation of the individual measurements. It is thus practicable to set up sampling programs so that most interpolated values of a variable will have standard deviations no greater than that of the measured values.

The result that interpolated values of a variable can be made more accurate than the observed values may at first seem a little surprising. This conclusion can best be understood from investigating the following sequence of conditions. First, consider the interpolation to be a procedure for taking a weighted average of observed values from the same location. This averaging process improves the accuracy of the interpolated value. Secondly, consider the observation locations to be somewhat separated but still sufficiently close together, so that the value of the variable changes only slightly over the complete interval. In this case the accuracy of the interpolated value is not as good as that for the case where the observations are all made at the same place, but it can still be better than that for a single observation. Thirdly, if observation locations become sufficiently far apart or are spaced such that the interpolation does not act as an averaging process, the accuracy of the interpolated value will become less than that for the observed values. The second condition can be assured by selecting the observation and interpolation locations such that the ratio, $\sigma_p{}^2/\sigma^2$, given by equation (15), is less than unity, and by spacing the observation locations sufficiently close together, so that the difference in the interpolation polynomials is so small compared to the standard deviation of the measurements that the right-hand term in equation (14) is negligible for all possible values of σ_r.

DEPTH ERRORS

Depth errors stem from two sources : first, errors in the actual determination of sampling depths; and second, errors because internal waves displace water with a given set of properties from its mean depth.

When the depth determination is the source of error, the errors at adjacent depths are most likely to be either the same or proportional to the depth. Under these circumstances, the interpolation formula yields the correct interpolated value for a depth which has the corresponding type of depth error. That is, if the sampling depths are in error by a fixed amount, the interpolated depth will be in error by the same amount. Likewise, if the sampling depths are in error by a factor, then the interpolated depth will be in error by the same factor. These depth errors, therefore, are not affected by interpolation but apply equally to the observed and interpolated data.

Since internal waves change only the actual vertical distribution of properties, the previous interpolation formulae will give the correct instantaneous values of a property at the interpolated depths. The error introduced by the internal waves is then the same, whether the value at a given depth is obtained by measurement or by interpolation.

APPLICATION TO ACTUAL STATION DATA

Temperature and salinity data from a typical oceanographic station (Station 1) are given in TABLE 2. For each interpolation depth, the values of the interpolation polynomials, the interpolation error, and the ratio of the square of the standard

Table 2a. Temperature data for Oceanographic Station No. 1

Obserevd Data		*Interpolation Data*					
Depth (m)	*Temp.* (°C)	*Depth* (m)	p_{123}	p_{234}	$\frac{1}{2}(p_{123} + p_{234})$	$\frac{1}{2}\lvert p_{123} - p_{234}\rvert$	σ_p^2/σ^2
99	9·67	100	9·63	9·62	9·62	0·00	0·9
124	8·69	150	8·21	8·22	8·21	0·00	1·0
149	8·22	200	7·83	7·76	7·80	0·02	0·8
183	8·01	250	7·12	7·13	7·12	0·00	0·7
238	7·26	300	6·63	6·63	6·63	0·00	1·0
284	6·78	400	5·88	5·89	5·88	0·00	0·9
396	5·90	500	5·45	5·45	5·45	0·00	0·8
483	5·54	600	4·82	4·75	4·79	0·02	0·7
564	5·06	700	4·23	4·28	4·26	0·02	0·6
658	4·39	800	4·04	4·06	4·05	0·01	0·9
753	4·15	1000	3·63	3·61	3·62	0·01	0·7
944	3·76	1200	3·14	3·15	3·14	0·00	1·5
1135	3·28	1500	2·67	2·66	2·66	0·00	1·7
1423	2·78	2000	2·01	2·02	2·01	0·00	0·9
1467	2·72	2500	1·75	1·75	1·75	0·00	0·8
1948	2·06	3000	1·60	1·61	1·60	0·00	0·8

deviation of the interpolation polynomial to the square of the standard deviation of the measurement, are shown. FIG. 1 consists of vertical plots of temperature and salinity which show both the observed and interpolated points. Inspection of the interpolation errors reveals that the sampling density was sufficient to portray adequately the vertical profiles of temperature and salinity. There are only five depths for temperature and two for salinity where the interpolation error is non-zero, and at these depths the interpolation errors are small compared to the measurement error. Except for the interpolations at 1200 and 1500 metres, the standard deviation of the interpolation polynomials is less than, or equal to, the standard deviation of the measurements themselves. At these depths, the standard deviation of the interpolated value is large, due to the fact that the sample depths of 1423 metres and 1467 metres were too closely spaced. This close spacing was due to the overlapping of two casts. Use of such unequal depth spacing is to be avoided in an interpolation programme. In some cases, due to the effects of internal waves, data from two casts do not join well at the overlapping depth. In such cases, it would be incorrect to use data from both casts to interpolate for intermediate depths, and it would be necessary to treat the data from each cast separately. However, in our particular case, with the two exceptions, the standard deviation of the interpolated values is everywhere less than, or equal to, the standard deviation of the measurement error.

Table 2b. Salinity data for Oceanographic Station No. 1

Observed Data		*Interpolation Data*					
Depth (m)	*Sal.* (‰)	*Depth* (m)	p_{123}	p_{234}	$\frac{1}{2}(p_{123}+p_{234})$	$\frac{1}{3}\lvert p_{123}-p_{234}\rvert$	σ_p^2/σ^2
99	33·53	100	33·54	33·54	33·54	0·00	0·9
124	33·66	150	33·76	33·77	33·76	0·00	1·0
149	33·76	200	34·02	33·98	34·00	0·01	0·8
183	33·96	250	34·04	34·04	34·04	0·00	0·7
238	34·03	300	34·08	34·07	34·08	0·00	1·0
284	34·07	400	34·10	34·11	34·10	0·00	0·9
396	34·10	500	34·22	34·22	34·22	0·00	0·8
483	34·20	600	34·27	34·25	34·26	0·01	0·7
564	34·26	700	34·27	34·28	34·28	0·00	0·6
658	34·25	800	34·34	34·34	34·34	0·00	0·9
753	34·31	1000	34·44	34·44	34·44	0·00	0·7
944	34·42	1200	34·51	34·50	34·50	0·00	1·5
1135	34·49	1500	34·55	34·55	34·55	0·00	1·7
1423	34·53	2000	34·62	34·62	34·62	0·00	0·9
1467	34·54	2500	34·65	34·65	34·65	0·00	0·8
1948	34·61	3000	34·67	34·67	34·67	0·00	0·8

It is seen from FIG. 1 that a unique curve can be drawn through the observed and interpolated values.

Temperature and salinity data from another station (Station 2) are presented in TABLE 3 and in FIG. 2. The interpolated values of temperature and salinity at 400 and 500 metres show large standard deviations for two reasons. First, the effect of measurement error is large. Because of the close spacing of sampling depths immediately above those depths and the rather large change in spacing of sampling depths in this range, σ_p^2/σ^2 is equal to 10 and 6 for these two depths. Secondly, in this depth range there is a rapid change of the gradient of the property with depth, and thus there is also a rather large effect due to interpolation errors. Two things are evidently required to improve the accuracy of the interpolated values for temperature and salinity. It is necessary to space more bottles in the depth range from 300 to 500 metres, and also to increase the spacing between sampling depths gradually, rather than abruptly. Again at 700 and 800 metres, the interpolation error for temperature becomes excessive compared to the observation error. In order to improve the accuracy of the interpolated temperature value, it is necessary to make another observation within this depth range. However, the change of salinity with depth through this depth range is sufficiently gradual to be adequately represented by the selected sampling depths.

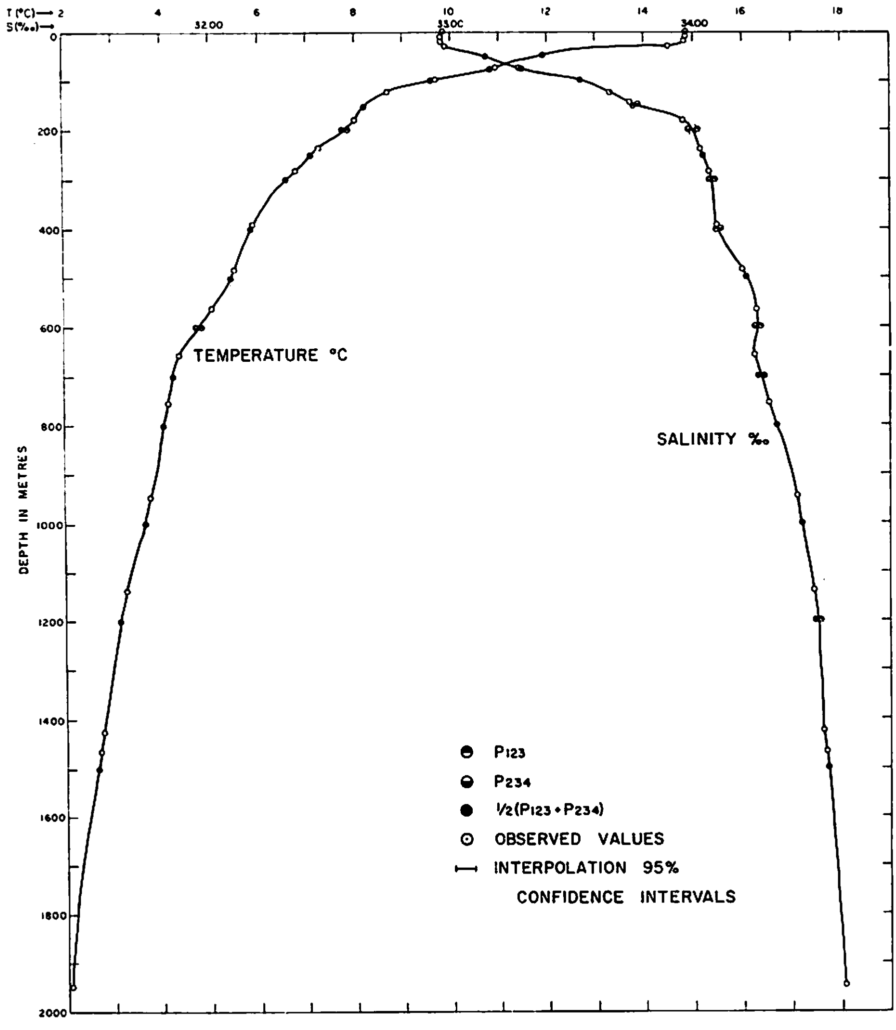

Fig. 1. Vertical distribution of temperature and salinity for Oceanographic Station 1.

These two examples illustrate that the errors of interpolated values arise from two separate causes. There is, first, the effect directly due to errors of measurement. This effect can be determined by the spacing of sampling depths and interpolated depths before any data are collected. The other cause of error is directly due to the interpolation. These errors cannot be determined ahead of time but depend upon the way the properties change with depth. After data have been collected, the interpolation error can be computed and used to determine whether the sampling density was sufficient to obtain a desired accuracy. Thus, for the data of Station 1, the density of sampling is sufficient to represent adequately the vertical profile. For Station 2, the data in the depth range from 300 to 500 metres are not sufficient for either temperature or salinity, while in the range from 600 to 1000 metres they are not adequate for temperature.

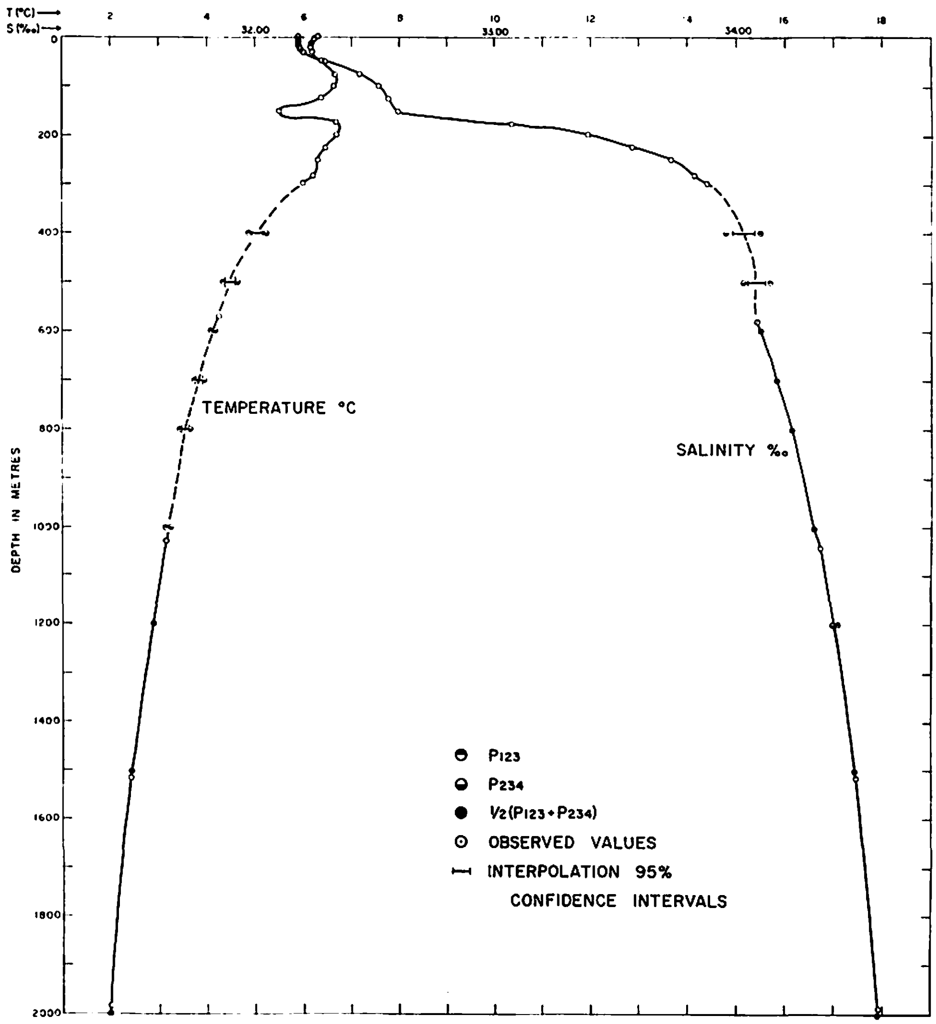

Fig. 2. Vertical distribution of temperature and salinity for Oceanographic Station 2.

APPLICATION TO OCEANOGRAPHIC SAMPLING PROGRAMMES

TABLE 4 shows the effects of measurement errors for selected spacings of sampling and interpolation depths. Inspection of this table shows that the best accuracy for interpolated values is obtained with equally distant spacing of the observation depths and with the depth of interpolation midway between the adjacent sampling depths. However, with the usual vertical distribution of properties, it is necessary to change the depth interval between samples, a procedure which increases the effects of measurement error. For conditions where the distance between adjacent depths is increased, the best results can be obtained by taking the interpolation depth to be from one-half to three-quarters of the distance between sampling depths in a direction toward the depths with larger observation spacing. With the aid of TABLE 4, any sampling programme can be designed to minimize the effect of measurement error on the interpolated values. The interpolation error, however, can be predicted only when a knowledge of the vertical distribution of the properties is previously available. Otherwise, in setting up a sampling programme, some vertical distribution of properties will have to be assumed. On the basis of this model distribution, a sampling programme will be made with the expectation that the interpolation error will not become too large. Only after the data are actually obtained, and interpolation computations completed, is it possible to assert whether the actual interpolations are within the prescribed limits of accuracy. This information should be obtained as early as possible in any oceanographic cruise, and the results utilized in determining the appropriate sample-spacing for the subsequent stations. In order to make this ship-board check, a plot of the data is helpful in determining the critical locations at which to compute manually the interpolation polynomials. In this manner, a sampling programme can be established which utilizes the minimum number of observed depths and yet obtains data within a prescribed accuracy.

Table 3a. Temperature data for Oceanographic Station No. 2

Observed Data		*Interpolation Data*					
Depth (m)	*Temp.* (°C)	*Depth* (m)	p_{123}	p_{234}	$\frac{1}{2}(p_{123}+p_{234})$	$\frac{1}{2}\lvert p_{123}-p_{234}\rvert$	σ_p^2/σ^2
200	6·69	250	6·31	6·31	6·31	0·00	1·0
225	6·46	300	5·98	5·98	5·98	0·00	1·0
250	6·31	400	4·87	5·24	5·06	0·12	10·0
284	6·21	500	4·32	4·62	4·47	0·10	6·0
300	5·98	600	4·11	4·18	4·14	0·02	1·0
568	4·26	700	3·71	3·93	3·82	0·07	0·8
1041	3·18	800	3·42	3·69	3·56	0·09	0·8
1515	2·42	1000	3·17	3·26	3·22	0·03	0·9

Table 3b. Salinity data for Oceanographic Station No. 2

Observed Date		*Interpolation Date*					
Depth (m)	*Sal.* (‰)	*Depth* (m)	p_{123}	p_{234}	$\frac{1}{2}(p_{123}+p_{234})$	$\frac{1}{2}\lvert p_{123}-p_{234}\rvert$	σ_p^2/σ^2
200	33·39	250	33·73	33·73	33·73	0·00	1·0
225	33·57	300	33·88	33·88	33·88	0·00	1·0
250	33·73	400	34·10	33·96	34·03	0·05	10·0
284	33·83	500	34·14	34·03	34·08	0·04	6·0
300	33·88	600	34·10	34·10	34·10	0·00	1·0
568	34·08	700	34·17	34·17	34·17	0·00	0·8
1041	34·35	800	34·23	34·23	34·23	0·00	0·8
1515	34·49	1000	34·33	34·33	34·33	0·00	0·9

It is immediately evident that missing data will have a very significant effect on the accuracy of interpolation. This effect is not due to this method of interpolation but is common to all methods and is shown in its true importance only by application of the preceding formulae. Every effort must be made to prevent occurrence of missing data in a sampling programme.

Table 4. Effect of measurement errors for selected spacings of sampling and interpolation depths

Depth Spacings				Error Analysis		
$x_2 - x_1$	$x_3 - x_2$	$x_4 - x_3$	$\frac{x_0 - x_2}{x_3 - x_2}$	σ_p^2/σ^2	σ_Δ^2/σ^2	$\Sigma/[\sigma^2(\bar{r}-r_0)]$
1	1	1	0	1·0	0·0	0·0
1	1	1	1/4	0·7	0·2	−0·3
1	1	1	1/2	0·6	0·3	0·0
1	1	1	3/4	0·7	0·2	0·3
1	1	1	1	1·0	0·0	0·0
1	2	2	0	1·0	0·0	0·0
1	2	2	1/4	0·9	0·4	− 0·7
1	2	2	1/2	0·8	0·7	− 0·5
1	2	2	3/4	0·8	0·4	0·1
1	2	2	1	1·0	0·0	0·0
1	2	3	0	1·0	0·0	0·0
1	2	3	1/4	0·9	0·3	− 0·7
1	2	3	1/2	0·8	0·6	− 0·6
1	2	3	3/4	0·7	0·3	− 0·1
1	2	3	1	1·0	0·0	0·0
1	1	2	0	1·0	0·0	0·0
1	1	2	1/4	0·7	0·1	− 0·3
1	1	2	1/2	0·6	0·2	− 0·2
1	1	2	3/4	0·7	0·1	0·1
1	1	2	1	1·0	0·0	0·0
2	2	3	0	1·0	0·0	0·0
2	2	3	1/4	0·7	0·1	− 0·3
2	2	3	1/2	0·6	0·2	− 0·1
2	2	3	3/4	0·7	0·1	0·2
2	2	3	1	1·0	0·0	0·0
2	3	3	0	1·0	0·0	0·0
2	3	3	1/4	0·8	0·3	− 0·5
2	3	3	1/2	0·7	0·5	− 0·2
2	3	3	3/4	0·8	0.3	0·2
2	3	3	1	1·0	0·0	0·0

OTHER APPLICATIONS OF INTERPOLATION FORMULAE

The examples used in this paper are all based upon depth interpolations of oceanographic data. However, in the presentation of oceanographic data by means of horizontal sections, the isopleths for any property are based essentially on interpolations between adjacent station data. The rules expressed above would apply equally well to these interpolations, and likewise, in the determination of the accuracy of the isopleths drawn to represent the observed data. It is desirable that this objective test be made on interpolations utilized in the drawing of these horizontal sections.

The analysis in this paper includes cases of one-dimensional interpolation only. Oceanographic data must often be interpolated in three or, when time is included, even four dimensions. An extension of this one-dimensional interpolation scheme would be appropriate for these three- and four-dimensional interpolations.

CONCLUSIONS

Error determinations are a necessary part of any series of experimental observations. In the oceanographic case where a series of discrete observations is used to represent a continuous distribution of properties, it is also required that the error be estimated for each point in the derived distribution. A method satisfying this requirement has been presented which estimates the error for each interpolated value obtained from a series of discrete observations. Its use permits a sampling programme to be planned to obtain the derived distribution of a variable to any desired accuracy.

Acknowledgements—The author is indebted to the members of the Eastern Pacific Oceanic Conference Committee on Machine Processing of Oceanographic data, whose many comments and suggestions on the problem of interpolation of oceanographic data have helped in the formation of the ideas presented in this paper. He is also indebted to the personnel of the Data Analysis Section of the Department of Oceanography, University of Washington, who have written the computer programme and carried out the computations.

This research was supported by the Office of Naval Research, Contract Nonr-477 (10), Project NR 083 012.

REFERENCES

BUCKINGHAM, R. A. (1957) *Numerical methods.* Pitman, London. 597 pp.

FOFONOFF, N. P. and TABATA, S. (1958) Programme for oceanographic computations and data processing on the electronic digital computer ALWAC III-E, DP-1. Oceanographic station data programme. *Mss. Rept., Fish. Res. Bd. Canad.* Pacific Oceanographic Group (Nanaimo, B.C.), Ser. No. 25, 38 pp. (Unpublished manuscript).

FROESE, C. (1960) Programmes for oceanographic computations and data processing on the electronic digital computer ALWAC III-E, DP-3. Oceanographic station data programme. *Mss. Rept. Fish. Res. Bd. Canad.* Pacific Oceanographic Group (Naniamo, B.C.), Ser. No. 73, 23 pp. (Unpublished manuscript).

KILMER, R. E. and DUXBURY, A. C. (1959) An analytical system for interpolation and calculation of dynamic heights. *Mss. Rept. Texas A & M, Ref.* No. 59–24T, 32 pp. (Unpublished manuscript).

STRICKLAND, J. D. H. (1958) Standard methods of seawater analysis. Vol. I. *Mss. Rept. Fish. Res. Bd. Canad.* Pacific Oceanographic Group (Nanaimo, B.C.), Ser. No. 18, 143 pp. (Unpublished manuscript).

TULLY, J. P. (1957) Some characteristics of sea water structure. *Proc. Eighth Pac. Sci. Cong.*, (Manila), 1953, 3, 643–661.

WOOSTER, W. S. and TAFT, B. A. (1958) On the reliability of field measurements of temperature and salinity in the ocean. *J. Mar. Res.*, **17**, 552–566.