FISHERIES RESEARCH BOARD OF CANADA

MANUSCRIPT REPORT SERIES
(OCEANOGRAPHIC and LIMNOLOGICAL)
No. 227

A Brief Study of the Accuracy of the Protected Deep-Sea Reversing Thermometer

by
Farrell M. Boyce

Pacific Oceanographic Group
Nanaimo, B.C.

December 31, 1966

Programmed
by
THE CANADIAN COMMITTEE ON OCEANOGRAPHY

1867 1967

FISHERIES RESEARCH BOARD OF CANADA

MANUSCRIPT REPORT SERIES

(OCEANOGRAPHIC and LIMNOLOGICAL)

No. 227

A Brief Study of the Accuracy of the Protected Deep-Sea Reversing Thermometer

by

Farrell M. Boyce

Pacific Oceanographic Group

Nanaimo, B.C.

December 31, 1966

1867 1967

Programmed

by

THE CANADIAN COMMITTEE ON OCEANOGRAPHY

FOREWARD

Although this paper was prepared some years ago, it is considered to be no less topical now than it was at the time of writing. The reversing thermometer continues to be the undisputed standard for precise measurement of temperature, as well as of depth, in the abyssal layers of the sea. Sensitive, complex, electronic instruments for recording continuously salinity and temperature with depth have recently been introduced; however, reversing thermometers are still necessary for in situ calibration of these instruments. Therefore, the proper handling and reading of reversing thermometers is as important now as it ever was, if not more so. It is to be noted that the unprotected reversing thermometer is not specifically mentioned in this report, but many of the results are as relevant to it as to the protected reversing thermometer.

The relatively rapid influx of new personnel into the marine sciences during recent years has often resulted in too-rapid and inadequate training of technical staff. The proper handling and reading of thermometers has sometimes been neglected during indoctrination into sea-going techniques. It is hoped that this report will point out some of the temperature errors that can arise from simple carelessness in reading reversing thermometers.

The author of this report makes no claim to have consulted all the available literature on accuracy and malfunctions of reversing thermometers. There are numerous papers on calibration and correction of reversing thermometers, as well as reports dealing with their malfunctions and with some of the possible remedies. However, we believe that the tests conducted by the author on precision and reproducibility in the reading of reversing thermometers are unique. From that point of view, it is considered desirable to reproduce this essay as a Manuscript Report in our Oceanographic and Limnological Series.

M. Waldichuk,
Oceanographer-in-Charge,
Pacific Oceanographic Group.

PREFACE

This essay is based on the author's summer work in 1959 at the Pacific Oceanographic Group and is supplemented by information drawn from the files of that Group. It was submitted to the Department of Engineering Physics of the Faculty of Applied Science, University of British Columbia, in partial fulfilment of the undergraduate requirements for the BA Sc. degree. The statistical work was formulated by Dr. N.P. Fofonoff, of the Pacific Oceanographic Group, and his assistance and advice greatly simplified the task of writing this essay.

Farrell M. Boyce

October 31, 1959

CONTENTS

APPENDIX A

APPENDIX B

ILLUSTRATIONS AND TABLES

A Brief Study of the Accuracy of the Protected Deep-Sea Reversing Thermometer

Introduction

The purpose of this paper is to describe the protected deep-sea reversing thermometer, its development and its operation, to estimate the human and physical factors affecting its accuracy, and to offer suggestions as to how its accuracy may be improved, or to indicate in which directions further research might be most profitable.

Temperature Structure in the Oceans

The physical properties of sea water in situ are described by the temperature and salinity. These two data are used to determine the in situ density of the water, which in turn leads to dynamical studies of the ocean.

The distribution of temperatures with depth is complicated. If one plots a graph showing temperature and depth simultaneously, two things become apparent at once. (Figure 1) . First, there is an upper layer in the ocean where the temperature-depth gradients are large. Second, there is a lower layer where the temperature-depth gradients are very small. It would appear from a first consideration that temperatures decreased steadily with depth, but this is not always the case. In the upper layer, 0 to 300 metres, there are marked seasonal variations in the temperatures, due largely to the changes in climatic conditions of the atmosphere above. In the lower layer, 300 metres to the bottom, the seasonal variations are correspondingly small. In the North Pacific Ocean the surface layer temperatures may range from 14°C to 3°C, while in the bottom layer the temperatures range from 3°C to a minimum of 1.5°C. However, when one considers the fact that the bottom layer extends from 300 metres to 6000 or 7000 metres it is at once apparent that the smallest measurable differences of temperature, ± .02°C with present-day equipment, will be of great importance when dynamical and thermodynamical studies are undertaken of so large a water mass.

In order to establish an overall picture of the temperature structure of the ocean at any one time, a large number of observations in different locations must be taken. These observations are taken from the deck of a ship, often in the most adverse conditions of weather and sea. The facts that many observations of a high degree of accuracy are required, and that these observations are obtained under poor conditions, necessitate a temperature-measuring instrument capable of measuring in situ temperatures that is sturdy, reliable, and above all, accurate.

Brief History of the Instruments Used to Determine the Temperature of the Oceans

The process of developing adequate thermometric apparatus for oceanographic work has been a long one. As early as the beginning of the eighteenth century efforts have been made to determine the temperature of the ocean depths. Two early methods are worthy of mention. The first method was to lower a large water-collecting apparatus over the side of a vessel and to trap a sample of water at the desired depth. This sample was hauled to the surface where its temperature was immediately measured. The defects of this method are obvious, yet with care and skill, utilizable results could be obtained for shallow waters. Many early studies of temperatures in the fresh-water lakes were carried out in this way. The second method involved the lowering of an ordinary liquid-in-glass thermometer which had a heavily insulated bulb. The thermomer was left at the desired depth a sufficient length of time for the thermometer to come to equilibrium against the effects of the insulation, then the thermometer was raised to the surface as quickly as possible and the reading taken. The temperatures of depth down to 1000 feet were measured in this fashion. The effects of pressure due to the great depths was a limiting factor in the early observations.

Temperature observations during the cruises of H.M.S. Porcupine and Lightning in the summers of 1868 and 1869, and during the voyage of H.M.S. Challenger during the years 1873 to 1876 were accomplished by means of maximum-minimum type thermometers. These thermometers had the advantage of being able to record an in situ temperature but they were subject to errors due to the increased pressure at great depths. They had a major disadvantage in that they recorded the minimum temperature of the water through which they passed and thus were incapable of detecting temperature inversions, which have subsequently been shown to exist as the rule rather than the exception in some bodies of water, notably the North Atlantic. An improvement on the early maximum-minimum type thermometers was the surrounding of the bulb by a glass shield, thus eliminating the pressure errors.

The Challenger expedition saw the first use of the reversing-type thermometer, the type in use today. The thermomètre à bascule was invented by a French oceanographer, Aimé and early models were manufactured by Negretti and Zambra of London, a firm still producing oceanographic thermometers. The first thermometers were of the outflow design and were operated by a slow 360° reversal accomplished by a propeller-driven rack. While they did overcome the serious drawback of the maximum-minimum thermometers by being able to detect temperature inversions, a comparison proved them to be less accurate. In 1876 Negretti and Zambra developed the double ended reversing thermometer which required an 180° reversal and which was built on the same principle as thermometers in use today.

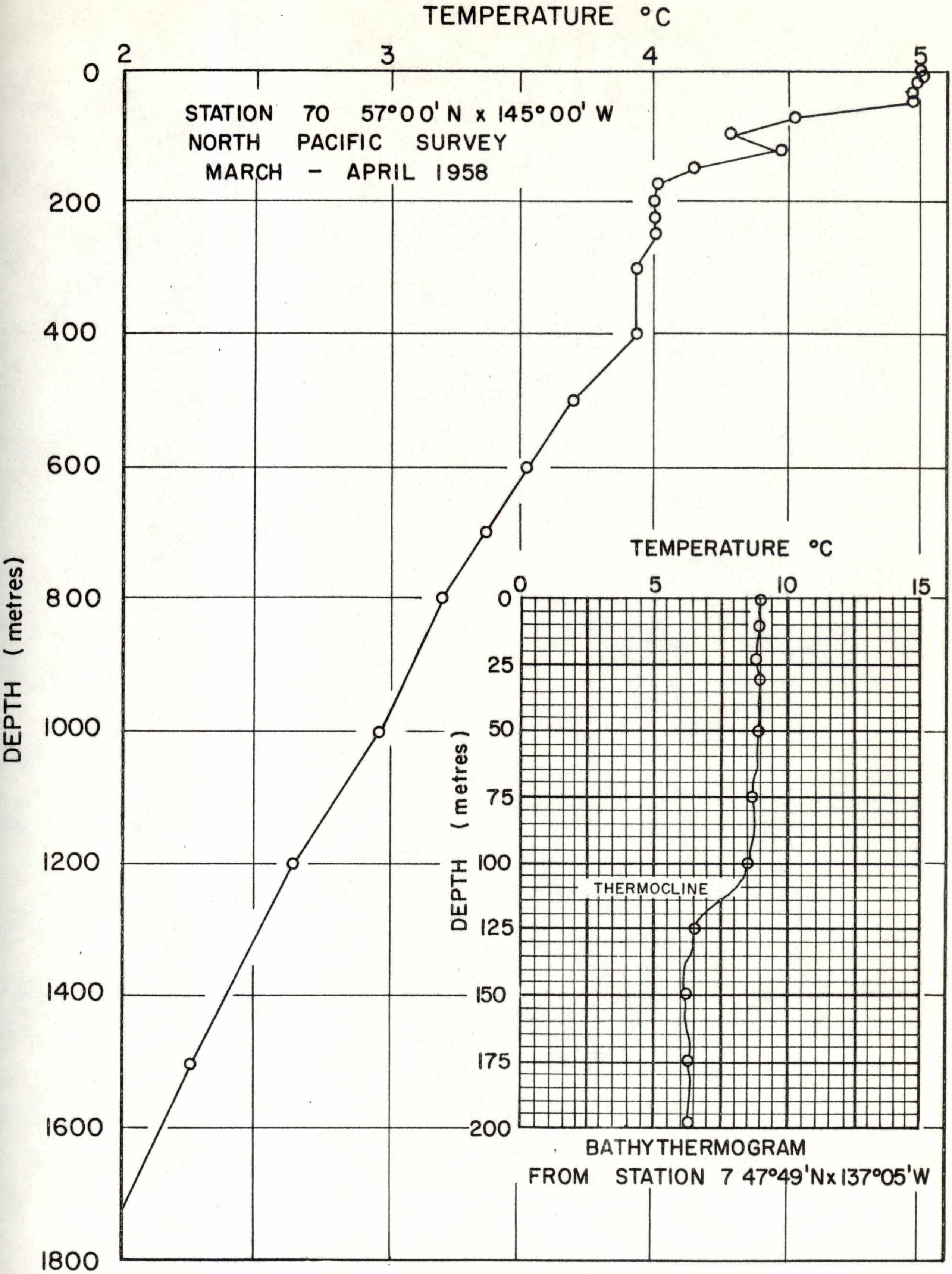


Figure 1. Typical Temperature Distributions in the North Pacific Ocean.

Work in thermometry, as far as oceanography is concerned, is far from stopping with the mercury-in-glass reversing thermometer. Thermometers with metallic elements, bourdon tubes, resistance thermometers, and thermocouples have all been adapted to the purpose with varying success. Of these, the most promising seem to be the resistance thermometer and the thermistor, to the extent that the Russians have used the former in a continuous temperature-depth recording device known as the Bathythermosonde. The Russians cannot claim to have pioneered this development, however; a resistance thermometer was in use aboard H.M.S. Challenger, and although it was cumbersome, its accuracy was comparable to the best of the thermometers in use at the time. The most reliable instrument for determining temperature still remains the reversing thermometer. The day of its total replacement is far from arriving, and until then it will be necessary to derive the utmost in accuracy from this instrument.

A Description of the Instrument, its Operation and Calibration

The protected reversing thermometer used in oceanographic work is basically a double ended outflow-type thermometer. Its principle of operation will be explained with reference to the accompanying diagram. (Figure 2). The thermometer is mounted in one of the racks of the reversing water bottles and sent down in the position shown (A to the bottom). As the thermometer comes to equilibrium, a volume of mercury, proportional to the temperature, is forced past the constriction (E) and into the loop (F) and the capillary tube (G). When the thermometer has reached an equilibrium (five minutes is usually sufficient time) the apparatus is reversed 180 degrees by the bottle mechanism so that end (B) is now at the bottom. During the reversal, the mercury column breaks at the constriction (E) and the mercury which had been forced past the constriction now flows to the bulb (H) at the opposite end. The cul-de-sac passage shown at the constriction aids in the accurate breaking-off process. The temperature is read off against the graduations on the capillary tube (G), which are placed at .1°C or .05°C intervals depending on the range. The loop (F) prevents any further mercury from draining into the capillary tube once the reversal has been made. The bulb (C) is surrounded by mercury (D) to provide a better conducting link with the water. The whole apparatus is encased in a glass shield (J), which protects the instrument from pressure effects due to depth. Since the temperatures below the surface of the water may be considerably different from the temperatures at which the thermometer is read, provision must be made for the expansion or contraction of the mercury and glass in the capillary tube. Alongside the capillary tube of the main thermometer is mounted an auxiliary thermometer of standard construction which gives the temperature of the surroundings. The reading of the main thermometer and the reading of the auxiliary thermometer are used to compute the difference between the main thermometer reading on the surface and the true temperature at depth. The following formula is often used as a good approximation:

Tw = true temperature

T = main reading

t = auxiliary reading

Vo = zero point volume in °C

1/a = linear expansion of mercury relative to that of glass

C = scale correction

$$T - Tw = \Delta T = \frac{(Vo + Tw)\ (t-T)}{a} - C$$

In actual practice, several methods are used to calculate this correction. One is by means of a specially constructed circular slide-rule. Another method, developed by Dr. J.P. Tully of the Pacific Oceanographic Group, entails a graphical representation of the equation on a sheet of paper called a nomograph. The nomograph takes into account the Vo and the scale corrections of the instrument and thus a separate nomograph is required for each thermometer. Tables of temperature corrections have also been drawn up.

Because of small constructional deformities in the capillary and the bulb, and because of the high degree of accuracy demanded, each thermometer requires extensive calibration. The instruments are subjected to a series of reversals, usually three to five, in a tank of water which can be maintained at even temperatures over the range of calibration. After each reversal, the thermometers undergoing calibration are compared with a standard thermometer accurate to ± .003°C. At each calibration point, the arithmetical average of the differences between the instruments under test and the standard thermometer is computed. This procedure is repeated at suitable intervals along the temperature range of the instrument. Zero-point volume and ice-point determinations are also made and the necessary corrections are added to the calibration chart.

Errors of Temperature Determination

Under ideal conditions, a good instrument will be sensitive to changes of ± .03°C in temperature. The sources of error in temperature determinations using the reversing thermometer are many and varied, but for convenience they will be classified into three basic types.

1. Functioning Errors

All errors due to the minor variations in the breaking of the mercury column for each reversal, and gross errors arising from defects, temporary or otherwise, in the thermometers will be called functioning errors.

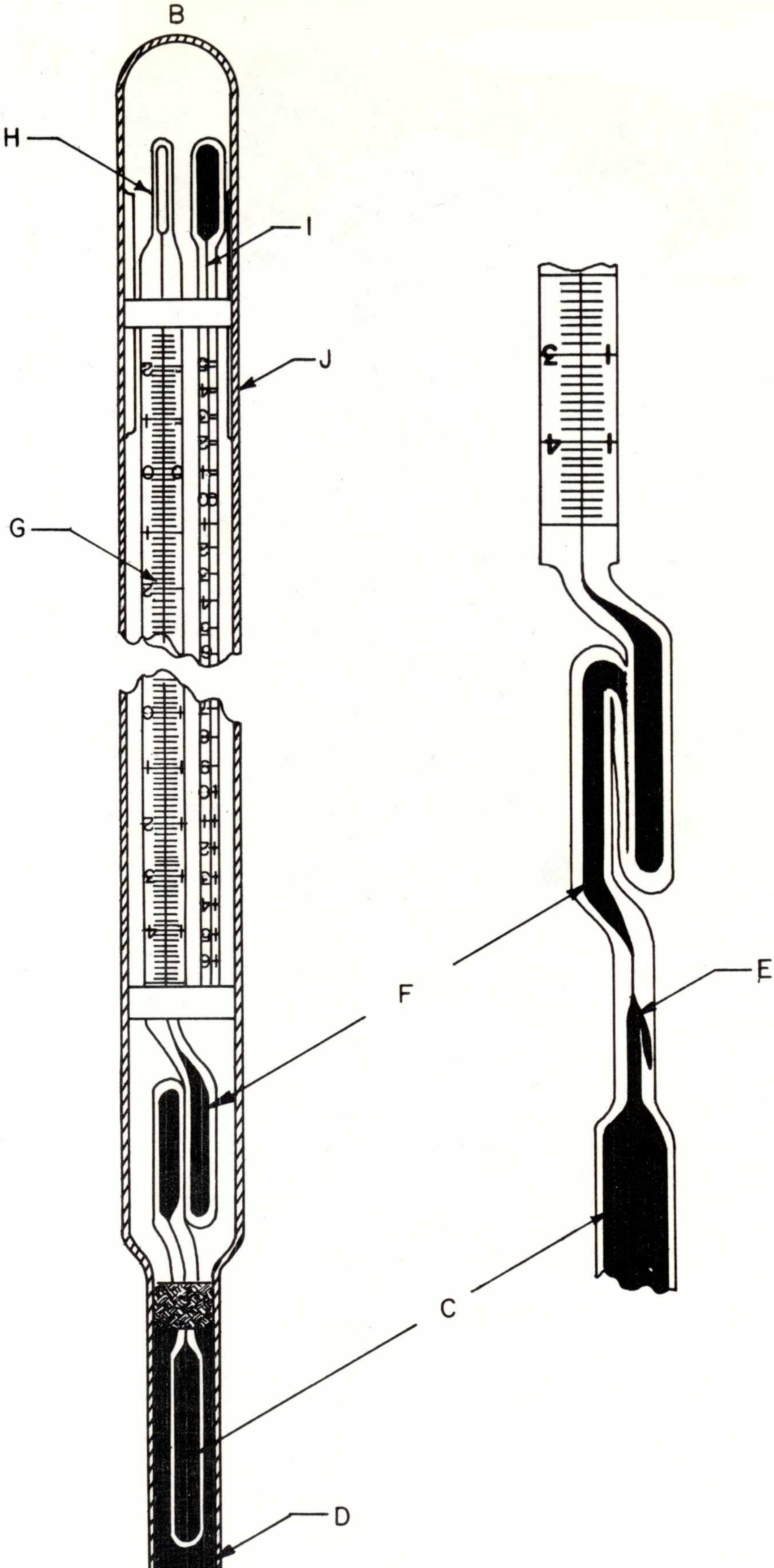


Figure 2. A typical Protected Reversing Thermometer.

2. Calibration Errors

Thermometers which are improperly calibrated or whose calibration has changed with age will show a systematic error. Another systematic error will arise if the wrong procedure is adopted in computing the correction or if a faulty nomograph is used. Errors from these sources will be classified as calibration errors.

3. Reading Errors

Errors which arise from interpolation between graduations, from parallax, and from the computation of the correction, will be known as reading errors.

Functioning Errors

Functioning errors may be split up into two types, normal functioning errors, and gross functioning errors.

(a) Normal Functioning Error

The conditions of reversal in the field are far from uniform. Individual differences among the bottles with respect to their reversing mechanisms, wire angles differing from the vertical, and the roll of the ship may all contribute to a greater or less degree to the functioning error, in that they may have an effect upon the speed of reversal, and hence upon the place of fracture in the mercury column. This functioning error will differ from thermometer to thermometer due, perhaps, to small individual differences in construction. Whitney (1957), optimistically I fear, places the functioning error in the order of ± .005°C.

(b) Gross Functioning Errors

Gross errors or malfunctions of three kinds have been observed. The most frequent malfunction observed on Pacific Oceanographic Group cruises was the empty-column type, wherein the thermometer was brought to the surface with the main capillary empty of mercury. It is not known exactly why this occurs, but it is probable that the mercury remains in the loop due to a defect in its formation. Among several thermometers, all of Japanese manufacture, this error was most persistent, and some of the thermometers were rendered useless.

The second most frequent defect was found to be that of a full main column. Reports of the Snellius Expedition of 1929-1930 (Hamaker, 1941) give this defect as their most frequent source of gross error. They state that it is caused by a separation between the mercury and the glass of the reservoir bulb. Again, this error seems to persist among several thermometers, but not to the extent of the empty-column defect. Japanese thermometers are the worst offenders.

The third malfunction was the occurrence of abnormally high readings of one of a pair of thermometers, which could not be attributed to the thermometers being reversed in the wrong place. This may have been caused by the mercury column breaking off at the wrong place in the thermometer due to an unusually narrow constriction at the neck of the reservoir. This would result in a reading several degrees too high. Errors of this type were not a frequent occurrence.

The Snellius Expedition reports the loss of 48 out of 6800 observations due to gross error. This figure would probably be higher if interpolation procedures were not adopted to fill in missing data. Out of 1000 separate observations taken on a Pacific Oceanographic Group cruise, 29 empty columns, 16 full columns, and 3 abnormally high readings occurred. In most cases, there was more than one thermometer at the depth so that these figures do not represent net losses. The actual loss of data is probably comparable to that of the Snellius Expedition.

Calibration Errors

As a rule, the thermometers are calibrated at equidistant intervals along their temperature range and any corrections required for readings which do not fall on these calibrated points are obtained by a linear interpolation of the calibration corrections. Referring to the graph of scale corrections plotted against position on the capillary, which has been taken from the Snellius reports (Figure 3A), it is easy to see how an error may occur and to form an estimate of its magnitude. In this case, the thermometer has been calibrated at intervals of one degree centigrade and the line of the graph represents the linearly interpolated correction values. Suppose instead, that the thermometer had been calibrated every two degrees and on the odd degrees. If this had been the case, the correction line would have followed the dashed line AC instead of the solid lines AB and BC. At the 6°C position, the calibration correction would have been in error by 0.01°C. No matter how carefully or how often the instruments are calibrated, this error will always occur, to say nothing of the error resulting from the original assumption that a linear approximation of the error between calibration points is sufficient. Whitney gives error from this source a value of ± .002°C while the Snellius Reports give it a value of ± .001°C. The magnitude of this error depends on the interval of calibration. Including an error of ± .005°C obtained in rounding off the scale corrections to the nearest .01°C, it seems reasonable to expect an error from this source of ± .008°C.

Investigations carried out while the Snellius Expedition was in progress showed that the zero point volume of (V_0) was in error for several of their

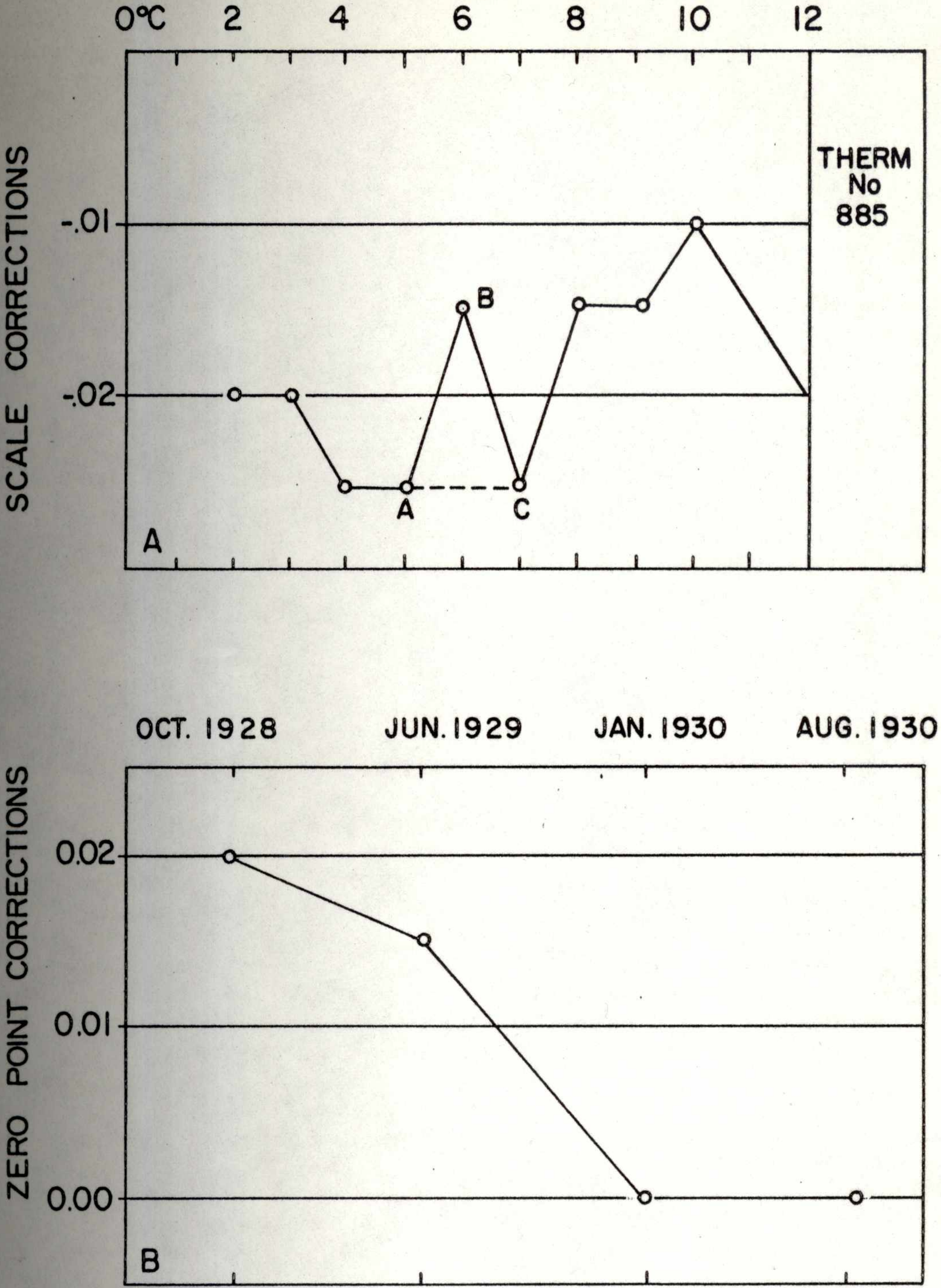


Figure 3A. A Calibration Chart.

3B. The Change in Zero-Point Correction with Time.

thermometers by an appreciable amount. This appears to have been due to a faulty calibration at the factory which manufactured the thermometers. While it is impossible to say whether the Vo of a thermometer is in error whithout an elaborate laboratory check, it is nevertheless a possible source of error. Careful calibration procedure should eliminate it altogether.

Further investigation on the Snellius Expedition showed that the zero, or ice-point of their thermometers was subject to a gradual change with time (Figure 3B). In all probability, this was due to the expansion or contraction of the reservoir bulb as stresses set up in the manufacture of the instruments gradually relieved themselves. Pre-war thermometers were aged for a period of two years before they were calibrated, but the increased demand for instruments has necessitated the dropping of this procedure. It seems very likely that new thermometers, bought since the war, will be subject to changes of this sort. This is probably the biggest contributing factor in the systematic differences observed between thermometers used in pairs.

Reading Errors

Reading errors, which may also be termed human errors, fall into three categories, interpolation errors, parallax errors, and computation errors.

1. Interpolation Errors.

It is standard practice to graduate the thermometers into intervals of 1/5, 1/10, or 1/20 of a degree and to read them to the nearest 1/100 of a degree. This requires a visual interpolation on the part of the observer. It has long been suspected that the error incurred in this interpolation is not a random error, but that it is an error characteristic of the observer, in that he tended to read high or low, to show a preference for certain numbers and an aversion to others, or to make more errors of interpolation at one place than another. A test was made of people's ability to interpolate between scale divisions in which the subjects were shown a series of cards upon which were drawn partially shaded oblongs representing the space between the graduations on a thermometer.[1] The cards were placed a distance away from the observers,corresponding to the degree of difficulty in reading actual thermometers. A series of 100 cards were shown to the subjects, they were asked to write down to the nearest 1/10 the amount the card was shaded, and then the figures they marked down were compared with a master sheet giving the correct values, and the results plotted up as a series of histograms (Figures 4A, 4B, 4C). Although 100 readings are a scant number with which to develop any statistics, there is, I believe, sufficient evidence to draw the following conclusions. First, there seems to be a

---

[1]See Appendix A for details.

pronounced tendency, varying among individuals, to read high, the remainder having either a uniform error distribution or else a negatively skewed one. Second, histograms showing the frequency with which a digit is chosen, indicate marked preferences for certain numbers. Again, these are subject to considerable individual variations, but in five out of the ten cases the distribution was bimodal with 2/10, 3/10, 7/10, and 8/10 being the preferred numbers, and 4/10 being the number most avoided. As might be expected, the majority of errors occurred in the middle of the ungraduated interval.

2. Parallax Errors

Since the actual graduations on the thermometers are placed behind the actual capillary tube and a small distance from it, and since the observer reads the instrument through a single magnifying lens with his eye roughly 10 centimeters from the mercury, it is reasonable to expect a certain amount of error due to parallax. This is illustrated in Figure 5. When this paper was planned, it was hoped to conduct a qualitative experiment similar to the one performed on interpolation error but time did not permit. However, from my personal experience regarding the roll of the ship, it does not seem unreasonable to assume that the error is a random one. In the Appendix A, there is a pseudo-statistical evaluation of the magnitude of parallax error, but of individual trends I can as yet say nothing.

3. Computation Errors

Errors of this type arise from the routine calculation of the temperature correction from the auxiliary thermometer reading. Just where these errors occur, it is hard to say. They may result from the rounding off of correction values to the nearest .01°C, or they may result from human errors in the actual computation of the correction. In either case, since they are defined to be distinct from the systematic errors discussed in a previous paragraph, it is probable that they are randomly distributed about a zero mean. With a little caution on the part of whoever calculates the corrections, errors of this sort can easily be kept to an acceptable minimum.

Statistical Formulation of Errors from Pacific Oceanographic Group Field Data[2]

To arrive at the error of temperature determinations in simple terms, two assumptions are made. First, that the three types of error, functioning, calibration, and reading error, are independent of one another; and second, that the standard deviations of functioning error and reading error are the same for all thermometers of a similar type.

---

[2]The statistical formulation is taken from the work of Dr. N.P. Fofonoff, then of Pacific Oceanographic Group, and is used, with his permission, in this paper. (Dr. Fofonoff is now with the Woods Hole Oceanographic Institution.)

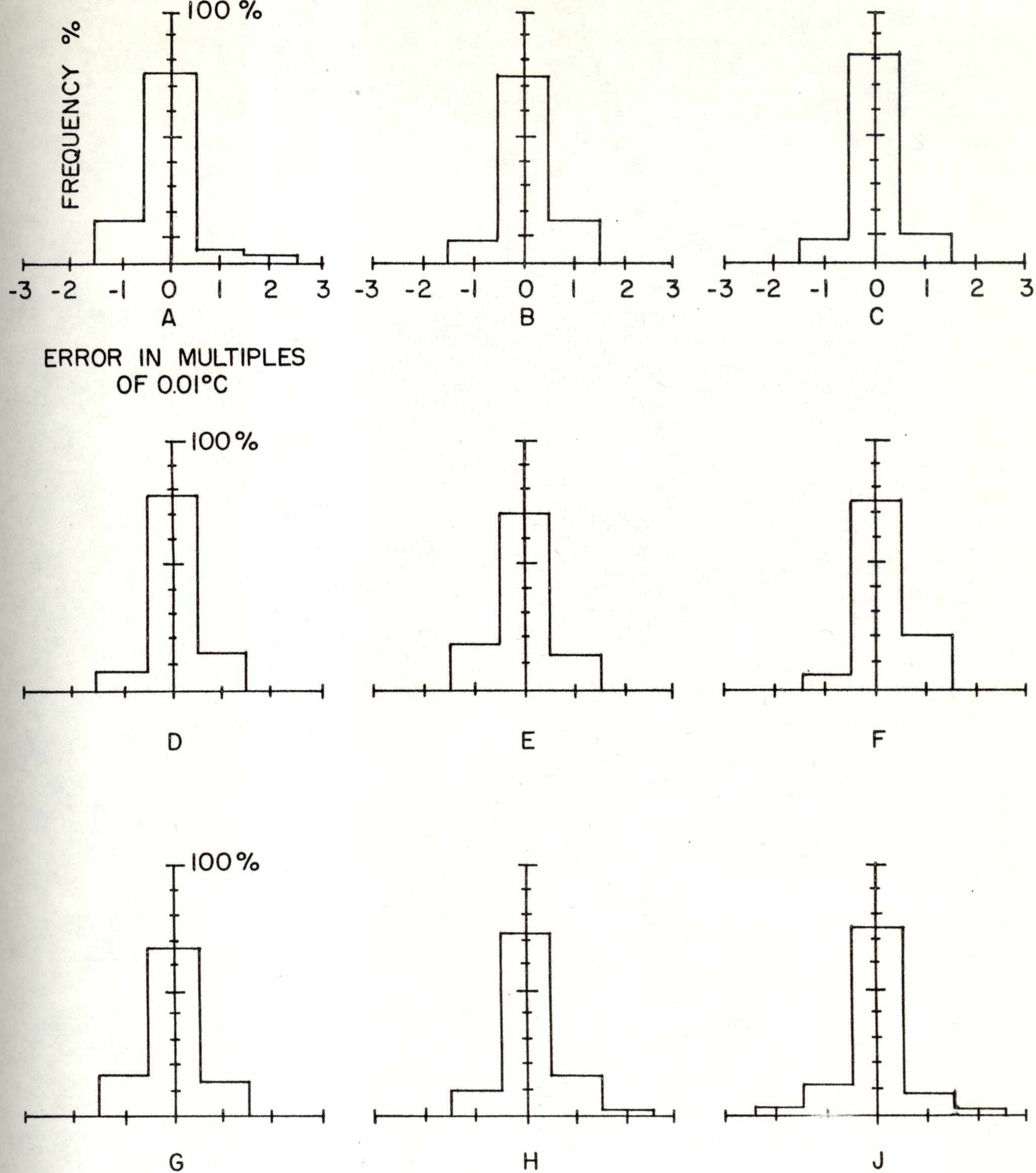


Figure 4A. Error-Distribution Histograms for the First Series of Tests.

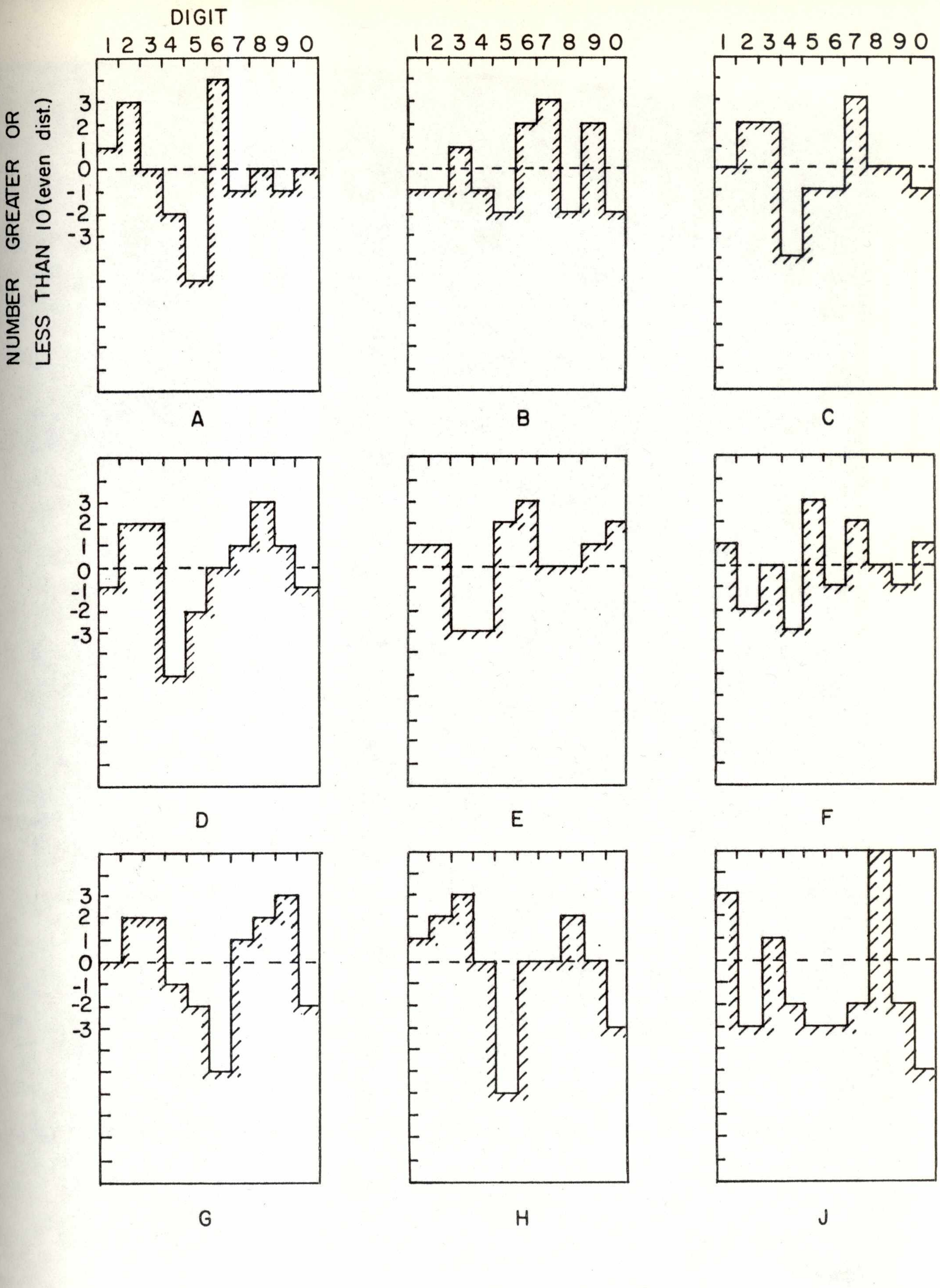


Figure 4B. Histograms Showing Digit Preferences.

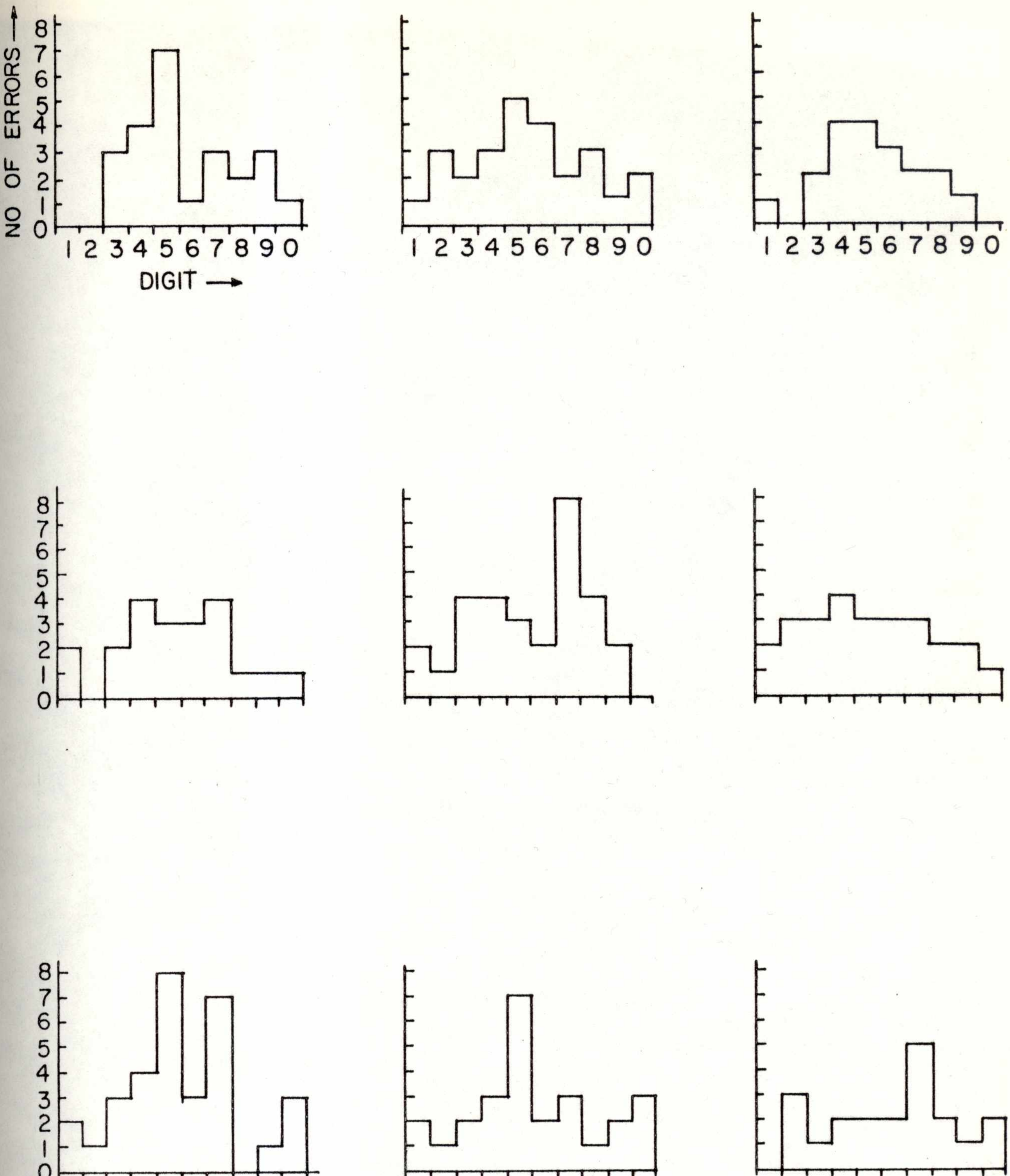
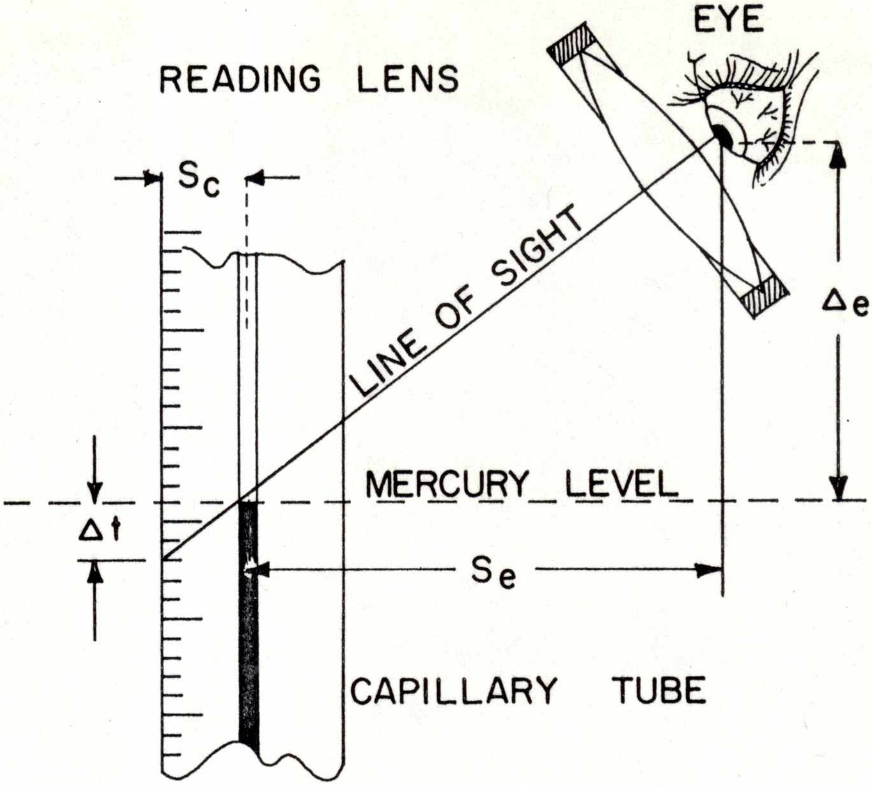


Figure 4C. Histograms Showing the Distribution of Error Among the Digits.



from similar triangles

$$\Delta_t = \frac{S_c}{S_e} \Delta_e$$

$$= .0125 \ \Delta e \ \text{inches}$$

where Δe in inches

$$.03'' = 0.1°C$$

$$\therefore \text{ if } \Delta e = 1'' \qquad \Delta t = \frac{.0125}{.03} = 0.42°C$$

$$\frac{\Delta_t}{\Delta_p} = 0.42\ °C/_{in}$$

Figure 5. The Geometry of Thermometer Reading.

Let the symbol σ_r indicate the standard deviation of reading a thermometer, the symbol σ_f indicate the standard deviation of the functioning of a thermometer, and the symbol $\Delta\vartheta_c$ indicate the mean difference between the readings of a pair of thermometers. At the 95% confidence level, assuming a normal distribution of error about a zero mean, the reading error of a single thermometer can be expressed as $\pm 1.96\,\sigma_r$, and the mean of n readings will then have an error of $\pm \frac{1.96\,\sigma_r}{\sqrt{n}}$.

The difference of two readings of the same thermometer should be normally distributed about a zero mean and have a standard deviation of $\sqrt{2}\,\sigma_r$. Thus σ_r can be found from field data wherein there is more than one reading of a single thermometer, which indeed is the usual practice. The formula used will then be $\sigma_r = \sqrt{\frac{1}{2n} \sum_{i=1}^{n} (T_{2i} - T_{1i})^2}$

The error σ_f and the mean difference between two thermometers $\Delta\vartheta_c$ can be obtained from a difference of two thermometers. This difference may be assumed to have a normal distribution with a mean of $\Delta\vartheta_c$ and a standard deviation σ, obtained from the formula $\sigma = \sqrt{2(\sigma_r^2 + \sigma_f^2)}$

Since σ_r can be determined by an independent method, it is possible to solve for σ_f by the formula $\sigma_f = \sqrt{\frac{1}{2}\sigma^2 - \sigma_r^2}$, where σ is given by $\sqrt{\frac{1}{m}\left[\sum_{i=1}^{m} (T_{A_i} - T_{B_i})^2\right] - \Delta\vartheta_c^2}$ and $\Delta\vartheta_c$ given by $\Delta\vartheta_c = \frac{1}{m} \sum_{i=1}^{m} (T_{A_i} - T_{B_i})$

The sum of the errors of temperature determination can be expressed in terms of σ_r and σ_f once the systematic errors between thermometers[3] are eliminated. The error of a mean taken from m thermometers read n times each is given by $\pm 1.96 \sqrt{\frac{\sigma_r^2}{nm} + \frac{\sigma_f^2}{m}} = \pm\Delta nm$ at the 95% confidence level.

1. The Difference of Two Means

The difference of two means computed from m thermometers read n times each should have a normal distribution about a mean difference and have a standard deviation of $\sqrt{2}$ times the standard deviation of a single mean. At the 95% confidence level the two means cannot be considered significantly different unless they differ by an amount greater than 1.96 times the standard deviation of the two means. Thus where T_1 and T_2 are the means of two separate determinations using m thermometers read n times, they are not considered significantly different unless $|T_1 - T_2| > \sqrt{2}\,\Delta nm$.

---

[3]See Appendix B

2. Calculation of Standard Deviation

Data collected in a recent North Pacific survey cruise during which 97 oceanographic stations were occupied in the Gulf of Alaska and along the Aleutian Archipelago were used to calculate the standard deviations. Each station consisted of 20 observed depths down to 1500 metres and thermometers were paired at five of these depths. Each thermometer was read twice with about 20 minutes between readings, the second reading being taken by a different observer. The data were considered to have been collected under typical field conditions.

The differences between corrected temperatures of the first and second readings, and between pairs of thermometers are given in Table 1, and are shown by the histograms in Figures 6 and 7. Some idea of the variability of the standard deviations can be obtained from Table 2.

An Overall Comparison of Thermometer Errors

A list of the contributing factors together with their size is given by Whitney (1957) and is displayed in Table 3. The reports of the *Snellius* expedition give the accuracy figure for a single temperature determination to be ± .011°C, and the German *Meteor* Expedition reports a value of ± .012°C, both figures being given for thermometers graduated in 1/10°C.[4] Assuming a standard deviation of a difference between two readings of 2 times the standard deviation of a single reading, at the 95% confidence level, two temperatures would not be considered significantly different unless they differed by 1.96 times the standard deviation of the difference. Applying these factors, temperature data from the *Snellius* expedition would not be considered significantly different unless they showed a difference from other data of .030°C, while temperatures from the *Meteor* data differing by less than .033°C would not be considered different. Evaluation of Pacific Oceanographic Group data by the method outlined in the preceding section places the significant difference figure at .04°C.

Conclusions Based on the Analysis of Pacific Oceanographic Field Data

Table 4 shows the relative accuracy of multiple readings of single and paired thermometers. The mean of a single reading of a pair of thermometers is more accurate than multiple readings of one thermometer. The mean of

---

[4] *Snellius* Reports. The standard error of a Single Temperature Observation if given by $S = \sqrt{\frac{\sum D^2}{\sum n}}$, where D denotes the difference between the indications of two paired thermometers and n is the number of pairs of observations.

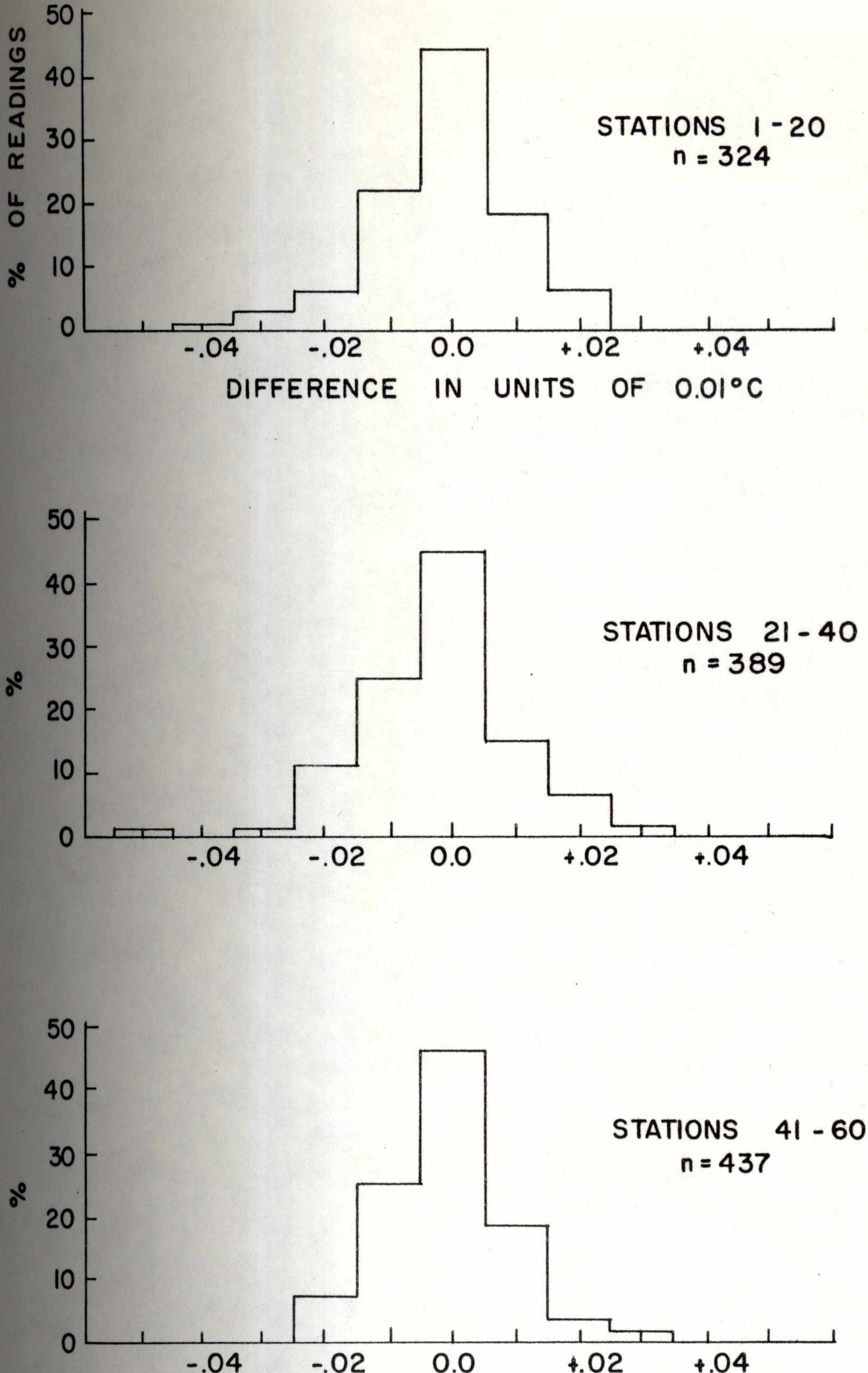


Figure 6. Histograms Showing the Distribution of Differences Between Two Readings of the Same Thermometer.

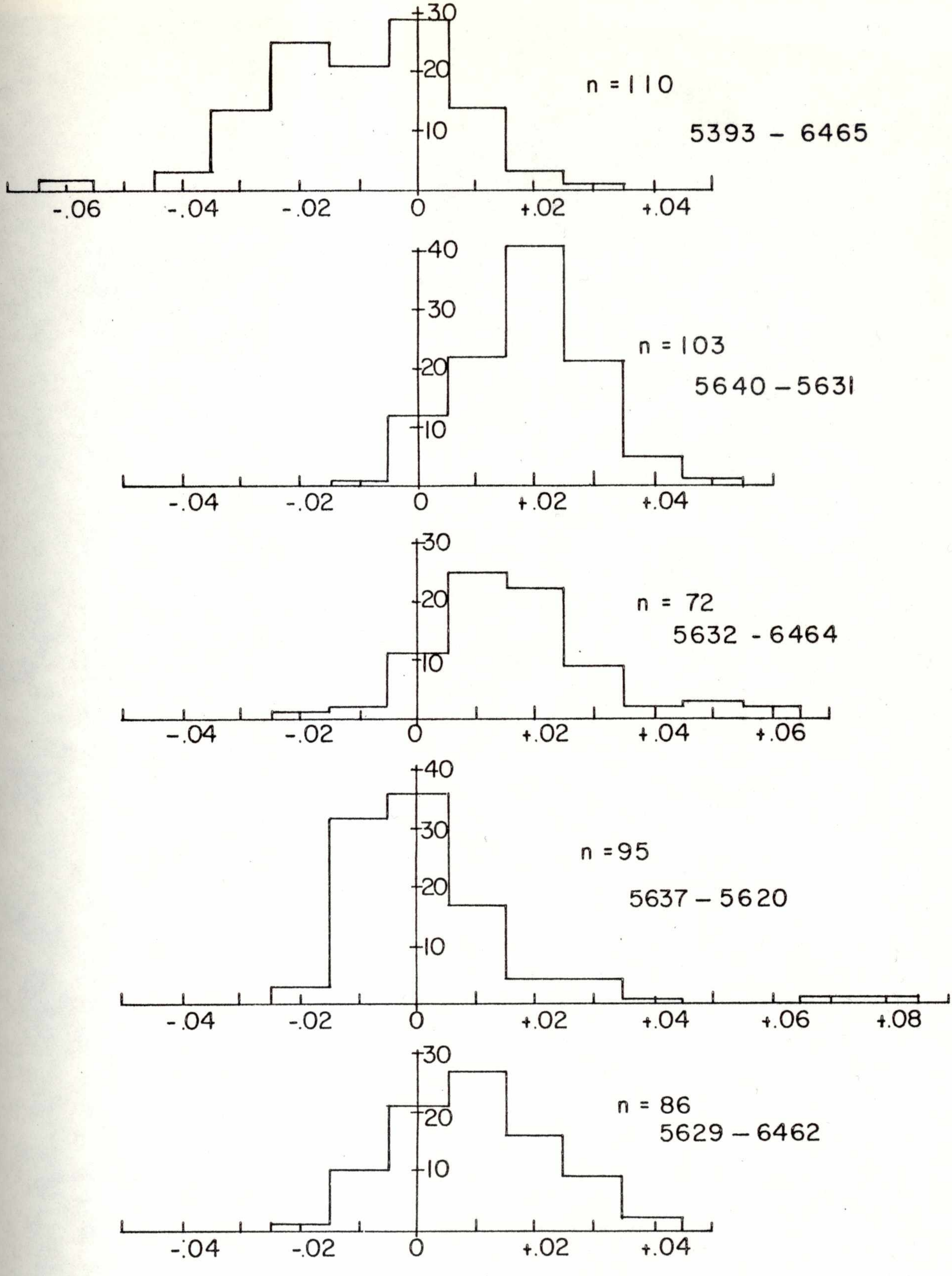


Figure 7. Histograms Showing the Distribution of the Differences Between Two Thermometers (including calibration and reading errors).

TABLE 1

FREQUENCY TABLE OF DIFFERENCES BETWEEN THERMOMETER READINGS

(A) Difference between first and second reading. in °C.

Stations	-04°	-03°	-02°	-01°	00°	01°	02°	03°	04°	05°	06°	Total
1 - 20	2	10	21	71	143	57	19	1				324
21 - 40	1	5	22	96	176	59	24	4				387
41 - 60		1	32	110	200	77	15	2				437
Total	3	16	75	277	519	193	58	7				1148

(B) Difference between thermometers. in °C

Thermometer No.	-04°	-03°	-02°	-01°	00°	01°	02°	03°	04°	05°	06°	Total
5393 - 6464	3	14	25	21	29	14	3	1				110
5640 - 5631				1	12	22	41	21	5	1		103
5632 - 6464			1	2	6	25	22	9	2	3	2	72
5637 - 5620			3	32	36	17	3	3	1			95
5629 - 6462			1	10	21	27	16	9	2			86
Total												466

TABLE 2

STANDARD DEVIATIONS

(A) Standard deviations of readings

Stations	$10^4\sum(T_2 - T_1)^2$	n	$10^4 \sigma_r^2$	$10^2 \sigma_r$
1 - 20	419	324	0.647	0.80
21 - 40	436	387	0.563	0.75
41 - 60	402	437	0.460	0.68
Total	1257	1148	0.547	0.74

(B) Standard deviations of differences between thermometers and calibration errors.

Thermometer No.	$10^4\sum(T_A - T_B)^2$	$10^2\sum(T_A - T_B)$	n	$\Delta\vartheta_c$	$10^4\sigma^2$	$10^2\sigma$
5393 - 6465	330	- 102	110	-0.01	2.14	1.46
5640 - 5631	481	189	103	0.02	1.33	1.15
5632 - 6464	379	127	72	0.02	2.21	1.49
5637 - 5620	116	- 2	95	0.00	1.22	1.10
5629 - 6462	218	82	86	0.01	1.64	1.28
Total	-	-	466	-	1.69	1.30

$10^4 \sigma_f^2$ 0.30

$10^2 \sigma_f$ 0.55 (from totals)

TABLE 3

COMPOSITION OF TOTAL THERMOMETER ERRORS

Error	Contribution to Total
Accuracy of standard thermometer used in calibration	0.003°C
Error in scale corrections after calibration	0.002°C
Reading error	0.002°C
Inherent functioning error	0.005°C
Interpolation error	0.002°C
Errors in rounding off scale corrections to 0.01°C	0.005°C
TOTAL - - - - - - - - - - - - - -	0.024°C
Snellius Expedition	0.030°C
Meteor Expedition (German)	0.033°C
Pacific Oceanographic Group	0.040°C

The first part of the table is from Whitney's (1957) work on the accuracy of thermometric depth determinations. Total errors in temperature determinations for several expeditions are listed below for comparison.

TABLE 4

PRECISION OF THE MEAN TEMPERATURE FOR
PROTECTED REVERSING THERMOMETERS.

$$\pm \Delta nm = \pm 1.96 \sqrt{\sigma_r^2/nm + \sigma_f^2/m}$$

and least significant difference between means $\pm \sqrt{2}\ \Delta nm$

Number of independent readings of each thermometer.	Number of thermometers 1	2	3
1	± .018°C	± .013°	± .010°
	.026	.018	.014
2	.015	.011	.009
	.021	.016	.013
2	.014	.010	.008
	.020	.014	.011
4	.013	.009	.007
	.018	.013	.010

two thermometers, read twice each, may be expected to differ from the true temperature in about 6% of the trials by an amount of ± .02°C, while a single reading of one thermometer may be expected to differ from the true temperature by the same amount in 26% of the trials.

The histograms in Figure 7 indicate improper calibration for some of the thermometers, in that the differences between paired thermometers exceed the possible reading errors. Furthermore, the distribution of these differences is skewed or even bimodal, indicating a high frequency of abnormal behaviour.

Data examined did not appear to be of uniform precision. In the data reports, no distinction has been made in the past between temperatures which are derived from the readings of a single thermometer and those which are the mean of a pair of thermometers. The degree of uncertainty existing between two separate determinations, ± .04°C, is too large for satisfactory deep-water measurements. It would appear that often insufficient attention is paid to the techniques used in the actual reading of the thermometers.

Recommendations

Wherever possible, thermometers should be used in pairs. From a practical standpoint, the limit, due to cost factors and the risk of loss at sea, would be two protected thermometers to a bottle.

A distinction should be made in the data records between data obtained from a single thermometer and data obtained from a pair of thermometers.

That all thermometers used in deep-water work should be carefully calibrated at frequent intervals, is a foregone conclusion, but attention should be paid to the temperature range in which calibration is particularly required. As previously stated, the range of temperature variation in the water below a depth of 300 metres is small, 1.5°C to 4°C. Thus it would be desirable to have the calibration intervals much closer together over this range, say at half-degree intervals.

A program should be conducted in the field, using data obtained under actual operating conditions, to eliminate the systematic differences between thermometers used in pairs. For an outline of this procedure refer to Appendix B.

A discussion of the conclusions and recommenations to be drawn from the reading-error experiments is found in Appendix A. Apart from these, it is recommended that all personnel who read the thermometers have a thorough understanding of the problems involved in accurate temperature determinations, and that they obtain sufficient practice in the fine art of reading a thermometer before they make observations on their own.

If these recommendations were to be adopted on a routine basis, it should be possible to reduce the uncertainty figure from ± .04°C to ± .02°C, for significant temperature differences. Considering the extremely small horizontal and vertical temperature gradients existing in the deep water, this would be an important and worthwhile increase in accuracy.

References and Works Consulted

Hamaker, H.C. 1941. The Oceanographic Instruments and the Accuracy of the Temperature Observations, The Snellius Expedition in the Eastern Part of the Netherlands East Indies 1929-30. Vol. II, Part 1, Chapt. I, pp. 1-56.

Teramoto, T. 1958. Depth Determination in Oceanographic Observations. Record of Oceanographic Works in Japan 1958, Vol. 4, No. 2 (New Series), pp. 170-185.

Thompson, C. Wyville. 1874. The Depths of the Sea, an Account of the General Results of the Dredging Cruises of H.M.S.S. "Porcupine" and "Lightning" During the Summers of 1868, 1869 and 1870. MacMillan and Co., London, 527 pp.

Tizard, T.H., H.N. Moseley, J.Y. Buchanan and John Murray. 1885. Report on the Scientific Results of the Exploring Voyage of H.M.S. "Challenger", 1873-76, Narrative of the Cruise - Vol. 1, First Part, pp. 83-120.

Tully, John P. 1937. A Graphical Method for Calculating the Corrections on the Deep-Sea Reversing Thermometers. Extrait du Journal du Conseil International pour l'Exploration de la Mer, Vol. XII, No. 1, pp. 39-44.

Whitney, G.G., Jr. 1957. Factors Affecting the Accuracy of Thermometric Depth Determinations. Extrait du Journal du Conseil pour l'Exploration de la Mer, Vol. XXII, No. 2, pp. 167-173.

APPENDIX A

Interpolation and Parallax Errors

1. Interpolation Errors - Experimental Outline

As a means of estimating the interpolation error in reading a protected thermometer, the following test was devised and ten employees of the Pacific Oceanographic Group were subjected to it.

A series of 100 cards were made similar to those in the illustration (Figure 8). On these cards were drawn oblongs of identical size and shape which were then partially shaded in black. The purpose of each oblong was to represent the interval between graduations on a thermometer scale, the shaded portion being the amount occupied by the mercury of the thermometer. The amount of shading on each card was measured to the nearest 1/10 of the portion of the whole oblong obscured. Care was taken to ensure that an even distribution of 1/10 intervals was obtained; in other words, there were ten of the hundred cards that were shaded 4/10, etc. It must be noted that the cards were not shaded to the even tenths; if they had been measured to the nearest hundredth, the hundredth figure would have shown a random distribution among the 10 digits.

The cards were then shown to the persons under test, who were told to estimate by eye the nearest tenth to which the oblongs were shaded. The cards were placed 15 feet from the observers, a distance corresponding to the angle subtended in the eye of the observers by actual thermometer graduations. After the cards had been shown, the written sheets bearing the subjects' estimates of the amounts shaded were compared to a master sheet. Three curves giving the results of the test were plotted for each person. The first was a histogram of the error distribution, the second a histogram showing the frequency with which one particular number was chosen, and the third, a histogram indicating the position and frequency of the errors along the ungraduated interval. The results of the first test are displayed in figures 4A, 4B and 4C.

A second test, conducted in a similar manner as the first, was made the next day. The following important changes were made. Fifty of the cards were altered by drawing a half-way line through the oblong, cutting the ungraduated interval in two. The remaining fifty were left unaltered, but 10 similar cards were made, each shaded to an exact tenth, and numbered so as to be clearly visible from a distance of 15 feet. Three tests were performed in this order. First, the fifty unchanged cards were shown in exactly the same manner as in the previous test. Second, the fifty cards bearing the half-way graduation were shown. Third, the first fifty cards were shown as before but the set of numbered cards, shaded to even tenths, were placed nearby to afford a comparison. The observers were instructed to use the standard cards as a means of gauging the shaded interval on the test cards.

The same three curves were drawn for the second series of tests as for the first. The results are shown in Figure 9. As a means of estimating the relative accuracy of each observer, the standard deviation of each error distribution histogram was calculated by the formula $\sigma^2 = \frac{1}{n} \sum \left[f(x_c - \bar{x})^2 \right]$. This figure was later used in connection with the parallax error.

Conclusions from the Interpolation Tests

The conclusions drawn from the first test are presented in the main body of the paper. The results from the first of the second series show much the same things as the first tests. A scrutiny of the curves in comparison with the earlier ones shows that marked changes among the individuals have occurred. In several of the cases, the skew of the error-frequency histogram changes. The digit-preference histogram is quite different for most people, and the accuracy, as shown by the standard deviation of error, has increased in every case, indicating a possible improvement with practice.

The results of the second test in the second series were perhaps the most striking. The test showed that the accuracy of the interpolation was not necessarily increased by dividing the interval in two. This is apparent from the percentage-accuracy figures and the standard-deviation figures for the test (Table 5). In 8 of the 9 results, the standard deviation is larger for the split-graduation test. The digit-frequency histogram appears to show no correlation with previous histograms. There seems to be a general preference for the number six. There is a preponderance of errors in the first half of the interval, corresponding to the intervals 1/10 to 5/10. There is some reason to believe that this test may not have been quite fair. It is possible that increased accuracy might have been obtained had the subjects been accustomed to the task of interpolating in a different interval.

The results of the third test of the second series show a marked increase in accuracy over the previous tests, with two exceptions. The skew of the error-frequency histogram, in most cases, is the same as the skew of the histogram for the first test of that series.

The great variability of the curves indicates an insufficiency of data, particularly from the viewpoint of statistical analysis. It may be that some of the difference between the accuracies of the tests with and without the split interval is due to an inequality in the two batches of cards. These two facts alone indicate the necessity for further and more thorough investigation before any really significant conclusions may be drawn.

2. Parallax Error

Figure 5 shows the geometry of thermometer reading. An approximate calculation shows that for every inch the eye is displaced above or below the actual level of the mercury in the capillary, an error of ± 0.042°C may be expected in the temperature figure. Bearing in mind the conditions under which the thermometers are read, the motion of the ship, and the fact that

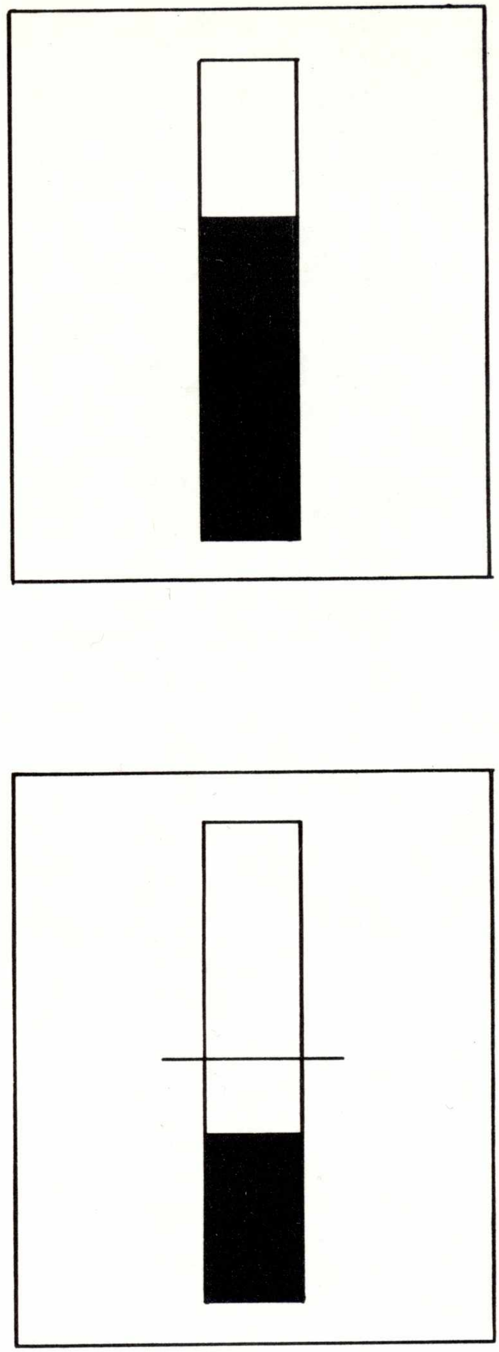


Figure 8. The Two Types of Interpolation Test Cards (Actual Size).

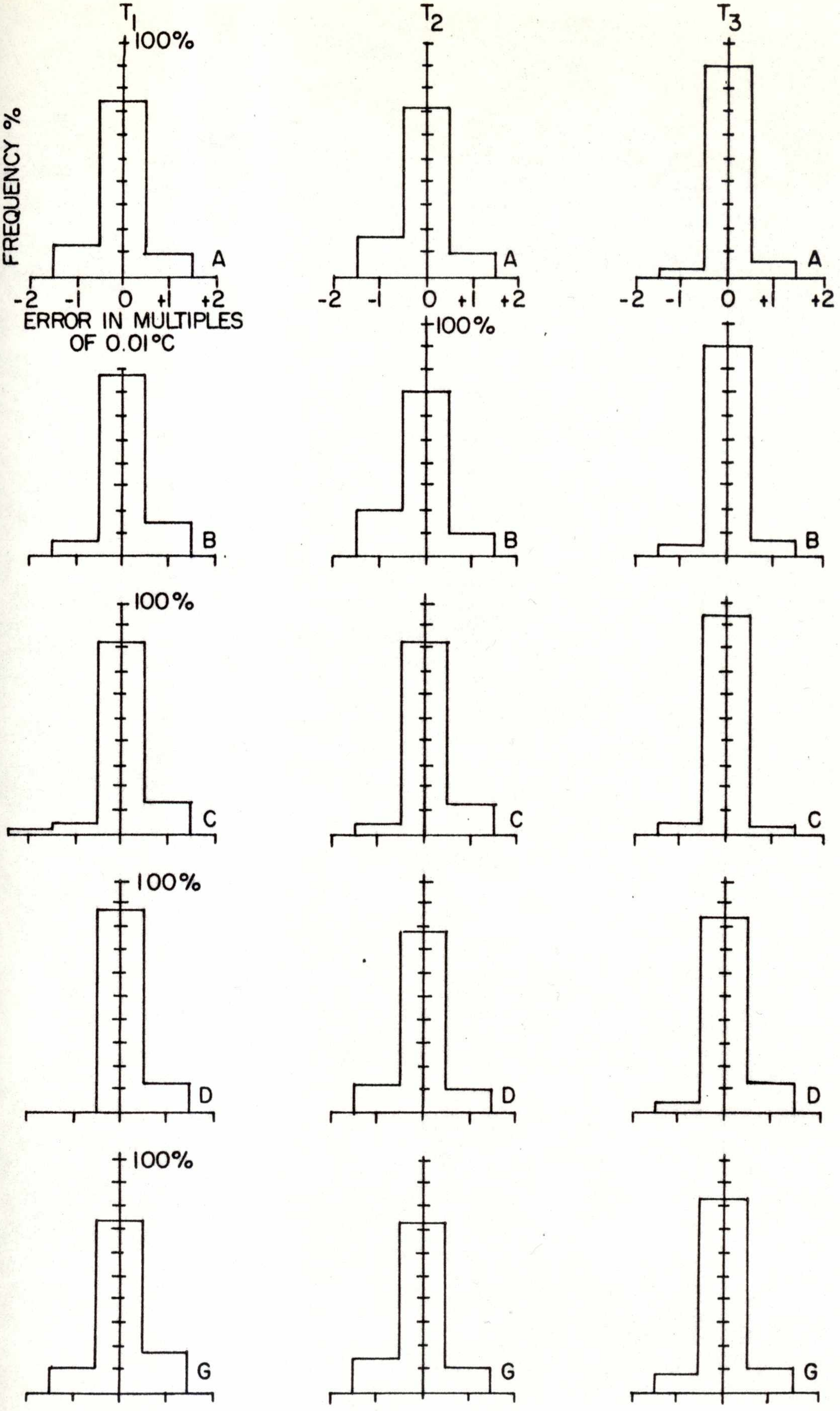


Figure 9. Error Distribution Histograms for the Second Series of Tests: T_1, as before; T_2, interval divided; T_3, with comparison cards of Individual Observers.

TABLE 5

STANDARD DEVIATIONS OF INDIVIDUAL OBSERVERS

Observer	T(100)	T(1)	T(2)	T(3)	Stns. 13-21	Stns. 47-55	Stns. 68-81
A	.58	.49	.53	.32	.89	.65	-
B	.51	.47	.58	.32	.93	.86	.81
C	.44	.42	.49	.25	.84	.74	.65
D	.47	.35	.47	.40	.83	.75	.80
E	.55	.40	.51	.45	-	-	-
F	.51	.47	.47	.45	-	-	-
G	.56	.53	.53	.42	.84	.76	.84
H	.55	.45	.53	.40	-	-	-
I	.59	.45	.51	.45	-	-	-

These Standard-Deviation figures are expressed as multiples of 0.01°C

TABLE 6

PARALLAX ERROR TABLE

Observer	p	p/i
A	.76	1.2
B	.87	1.4
C	.74	1.4
D	.79	1.3
G	.81	1.0

σp is Parallax error expressed in multiples of 0.01°C. $\sigma p/\sigma i$ is the ratio of interpolation error to parallax error.

the mounting of the bottle racks requires a stooping in order to read the instruments, it is not inconceivable that an error of this magnitude could enter into the reading. Although the validity of the following method is open to question, it will serve to give a rough idea of the magnitude of the error attributed to parallax.

During a recent Pacific Oceanographic cruise, the thermometers were each read four times at the stations by four individuals. If the mean of the four readings is assumed to be the true temperature, then the difference from the mean for individual readings may be taken as a measure of their inaccuracy, the difference being interpreted as the total error in reading the thermometer. From the cruise data, a series of histograms was drawn up (Figure 10), similar to the error-frequency curves drawn for the interpolation tests shown in Figure 9. A separate histogram was drawn for three legs of the cruise and the standard deviations were calculated for each histogram in the manner outlined previously. It was assumed, with respect to this calculation, that the error in the standard deviation caused by taking zero error as the mean value would be small. It will be noted that the histograms show a variation among themselves as to their skew and standard deviation. Accordingly, an average value of the standard deviation was taken. This value, $\bar{\sigma}$ is intended to represent the average standard deviation of error in thermometer reading for each individual. Using the identity $\bar{\sigma}^2 = \sigma_i^2 + \sigma_p^2$ where σ_i^2 and σ_p^2 are the partial variances of a total variance $\bar{\sigma}^2$, the partial error attributed to parallax was computed. The partial variance for interpolation error σ_i^2 was taken to be the value for the standard deviation from the first of the series of interpolation tests. Thus, knowing $\bar{\sigma}$ and σ_i, σ_p may be found from the formula $\sigma_p = \sqrt{\bar{\sigma}^2 - \sigma_i^2}$
Table 6 lists the standard deviation of error caused by parallax. This shows the parallax error to be comparable to or greater than the interpolation error. Nothing, however, can be said about the skew tendencies for parallax.

Conclusions and Recommendations

As a means of estimating the reading between graduations of a thermometer, the visual comparison method as employed in the final interpolation test is the most accurate. If all the thermometers were of uniform dimensions with respect to the graduations interval, an optical device might be employed in the reading lens. This, for the moment, is impractical. However, it would perhaps be worthwhile to draw up a series of small cards similar to the one shown in Figure 8 and to mount them on the bottle racks so that they may be referred to during the reading of the thermometers.

It appears that a person's accuracy improves with practice. This is illustrated by the series of histograms drawn from the tests, and the histograms shown in Figure 5, which show a diminishing of reading error as the cruise progressed. This may also be seen, to a lesser extent, in the histograms drawn for the estimation of parallax error, the histograms being drawn for chronologically arranged data.

Since optical devices are not, for the present, to be considered, the parallax error could be reduced by two ways. First, the bottle racks could be mounted so that the thermometers would be close to the eye-level of a standing observer. Second, if the observer were to adopt the practice of locating the level of the mercury and marking it with a finger, he would be more likely to place his eye at the correct level for reading the thermometer.

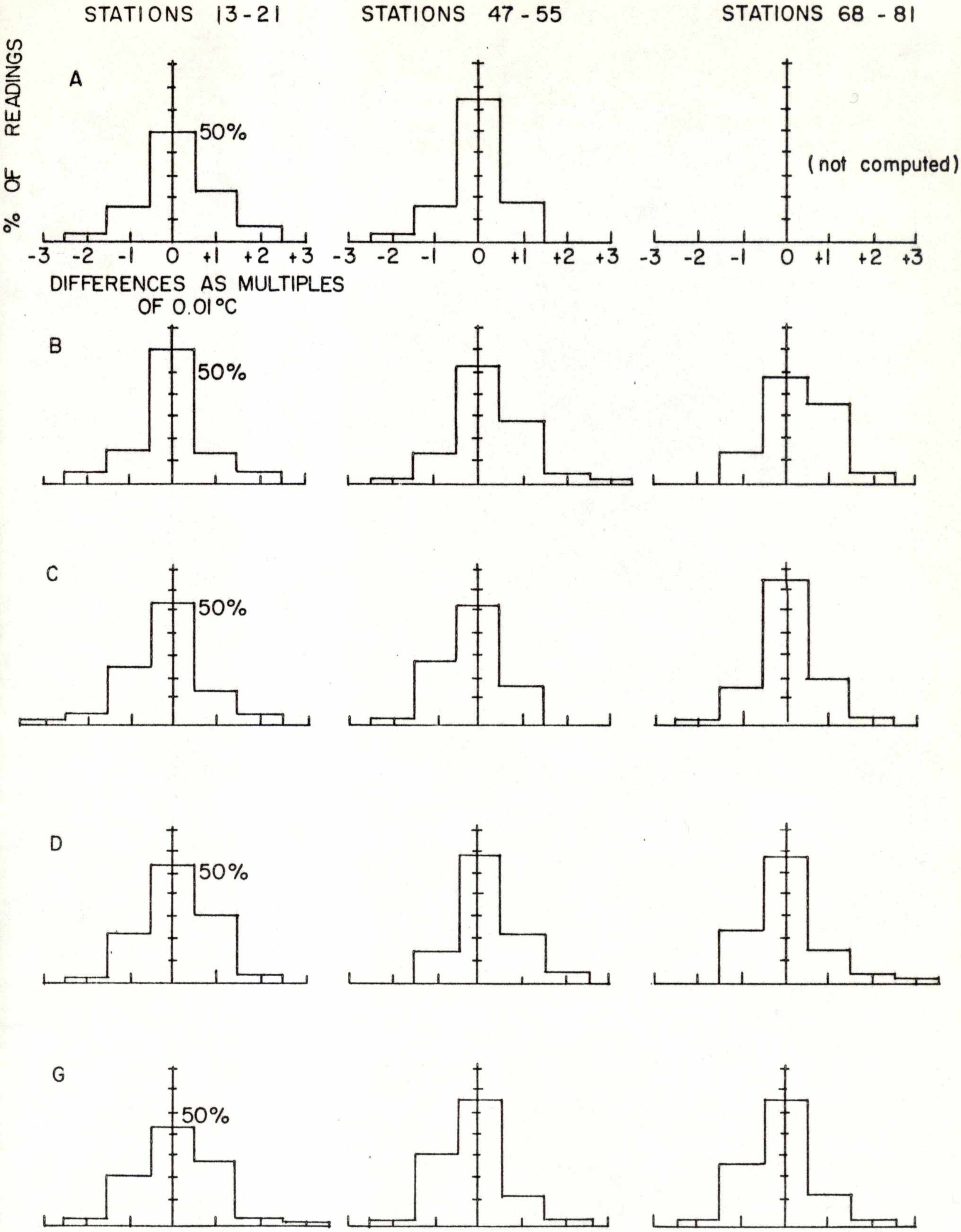


Figure 10. Histograms Showing the Frequency Distribution of the Difference Between Individual Readings and the Mean of Four Readings.

APPENDIX B

The Computation of Thermometer Corrections from Field Data

(Courtesy of Dr. N.P. Fofonoff)

As a means of eliminating the calibration errors between pairs of thermometers, the following method has been devised by Dr. N.P. Fofonoff. The thermometers used in deep water work are paired for at least ten reversals; then the pairs are interchanged to form a new set of pairs, which again are operated for a series of reversals. In this fashion each thermometer is compared with as many others as possible. Among the thermometers are a number of carefully calibrated instruments which are assumed to give a true reading. Four thermometers for this purpose were calibrated by The National Research Council of Canada.

The averaged differences between thermometers are set in a matrix form.

	C_1	C_2	....	C_m
r_1				
r_2		$\Delta_{ij} = \bar{T}r_i - TC_j$		
⋮				
r_n				

The problem is to find a set of corrections δr_i for the row thermometers and δC_j for the column thermometers that will give the best estimate of the true temperature.

If there were no statistical variation in the differences between thermometers, the elements of the matrix Δ_{ij} would be given by $\delta C_j - \delta r_i$

This, however, is not the case.

Selecting a set of indices, we arrange that the sum of the squares of the differences between the two matrices is minimized. Let A be taken so that $A = \sum_i \sum_j [\Delta_{ij} - (\delta C_j - \delta r_i)]^2$. A will be minimized if

$$\frac{\partial A}{\partial \delta C_j} = 2\sum_i (\Delta_{ij} - \delta C_j + \delta r_i) = 0 \quad \text{and} \quad \frac{\partial A}{\partial \delta r_i} = 2\sum_j (\Delta_{ij} - \delta C_j + \delta r_i) = 0$$

Therefore we must have

$$\delta r_i = -\frac{1}{m}\sum_j \Delta_{ij} + \frac{1}{m}\sum_j \delta C_j, \quad \delta C_j = \frac{1}{n}\sum_i \Delta_{ij} + \frac{1}{n}\sum_i \delta r_i$$

Solving for δr_i, we get $\delta r_i = -\frac{1}{m}\sum_j \Delta_{ij} + \frac{1}{m}\sum_j\left(\frac{1}{n}\sum_i \Delta_{ij} + \frac{1}{n}\sum_i \delta r_i\right)$

or $\delta r_i = -\frac{1}{m}\sum_j \Delta_{ij} + \frac{1}{nm}\sum_i\sum_j \Delta_{ij} + \frac{1}{n}\sum_i \delta r_i$. Similarly,

$$\delta C_j - \frac{1}{n}\sum_j \delta C_j = \frac{1}{n}\sum_i \Delta_{ij} - \frac{1}{mn}\sum_i\sum_j \Delta_{ij}$$

These equations do not determine the corrections explicitly, but involve an arbitrary constant ϵ. Thus, we can take

$$\delta r_i = \epsilon - \frac{1}{n} \sum_j \Delta_{ij} + \frac{1}{mn} \sum_i \sum_j \Delta_{ij}$$

$$\delta c_j = \epsilon + \frac{1}{n} \sum_i \Delta_{ij} - \frac{1}{mn} \sum_i \sum_j \Delta_{ij}$$

The constant ϵ must be determined by comparison with calibrated standard thermometers. For example, if we have a few, say S calibrated row thermometers, we can choose ϵ so that

$$\sum_S \delta r_i = S\epsilon + \sum_S \left(\frac{1}{mn} \sum_i \sum_j \Delta_{ij} - \frac{1}{m} \sum_j \Delta_{ij} \right) = 0$$

Solving for ϵ we get

$$\epsilon = -\frac{1}{S} \left[\sum_S \left(\frac{1}{mn} \sum_i \sum_j \Delta_{ij} - \frac{1}{m} \sum_j \Delta_{ij} \right) \right]$$

This method assumes, of course, that the complete matrix is available. However, no practical difficulties arise in applying it to an incomplete matrix.