NOAA Technical Memorandum ERL BOMAP-2

U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
Environmental Research Laboratories

N132KI

A Statistical Data Plan for BOMEX

THEODORE W. HORNER

The
BOMAP
Office

ROCKVILLE, MD.

December 1970

NOAA TECHNICAL MEMORANDA

Environmental Research Laboratories, BOMAP Series

The Barbados Oceanographic and Meteorological Analysis Project (BOMAP) was established in the Environmental Research Laboratories to coordinate the reduction of data collected during the Barbados Oceanographic and Meteorological Experiment (BOMEX) of May 3 to July 28, 1969; to complete analysis of the BOMEX Core Experiment (the Sea-Air Interaction Program of the experiment); and to provide a central contact point for BOMEX information exchange and publication of BOMEX results.

NOAA Technical Memoranda in the Environmental Research Laboratories BOMAP Series document methods, procedures, and techniques used to collect, analyze, and evaluate BOMEX data. They facilitate rapid dissemination of information that is preliminary in nature and subject to formal publication elsewhere at a later date.

Papers listed below are [illegible] from the National Technical Information Service, U.S. Department [illegible] Commerce, Sills Bldg., 5285 Port Royal Road, Springfield, Va. 22151. [illegible] $3.00 paper cover; $0.65 microfiche.

ERL BOMAP 1 High-Level Cloud Photography Inventory, BOMEX Period IV, July 11-28, 1969. Vance A. Myers, September 1970. (PB-194 661)

U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
Environmental Research Laboratories

NOAA Technical Memorandum ERL BOMAP 2

A STATISTICAL DATA PLAN FOR BOMEX

Theodore W. Horner
Booz-Allen Applied Research Inc.
Washington, D.C.

Final Report prepared for
Barbados Oceanographic and Meteorological
Analysis Project (BOMAP)
Contract No. E-192-69(N), Task 4 July 5, 1970

The BOMAP Office
Rockville, Maryland
December 1970

UDC 551.465.7:551.511:519.24/.25(083.3)(083.9)

551.46	Oceanology
.465.7	Interaction of the sea with the environment
551.5	Influence of the sea on weather
.511	Mechanical and thermodynamical aspects
519.2	Mathematical statistics
.24	Analysis and interpretation of statistical material
.25	Statistical regularity. Errors of observations
(083.3)	Equations
(083.9)	Plans

TABLE OF CONTENTS

INDEX OF FIGURES

INDEX OF TABLES

SUMMARY

As part of the United States contribution to the Global Atmospheric Research Project (GARP), a large-scale meteorological experiment, known as the Barbados Oceanographic and Meteorological Experiment (BOMEX), was conducted near the island of Barbados in the Caribbean from May through July of 1969. The primary goal of BOMEX was to provide data for studying the sea-air interaction that drives atmospheric circulation and global weather systems. To accomplish this goal, thousands of hours of data tape and other records were collected. Twelve ships and 29 planes were utilized. More than 1,500 personnel from government agencies, private industry, and universities participated in the project. The processing and coordination of BOMEX data were the responsibility of the Barbados Oceanographic Meteorological Analysis Project (BOMAP), a project established within the Research Laboratories of the Environmental Science Service Administration.

The reduction of the BOMEX data to a form useful for storage, scientific computation and interpretation presented many statistical problems. The object of the subject contract was to provide a methodology for estimating the terms of the budget equations and for assessing the precision of the estimates. The budget equations were equations for the flux of energy from the ocean to the atmosphere into and out of the BOMEX box (an imaginary box approximately 500 kilometers square and 500 millibars of pressure differential high in an ocean area near Barbados).

I. INTRODUCTION

The Barbados Oceanographic and Meteorological Experiment (Project BOMEX) was the first of a series of large-scale research projects planned by many nations throughout the world under the Global Atmospheric Research Project (GARP). The primary goal of the BOMEX project was to provide data for studying the sea-air interaction that drives atmospheric circulation and global weather systems. The project was conducted from May through July of 1969 and produced substantial quantities of data that were recorded manually, on magnetic tapes, and on charts. The reduction of the raw data to a form useful for scientific investigation and for storage in scientific archives is the responsibility of the Barbados Oceanographic and Meteorological Analysis Project (BOMAP), a project established within the Research Laboratories of the Environmental Science Services Administration.

1. DATA REDUCTION AND ANALYSIS

In the reduction of the data it has been important to consider not only the meteorological theory underlying the BOMEX experiment but also the statistical properties of the data and the mechanisms and formats used in assembling the raw data. The data-reduction process presents problems with respect to filtering, averaging, and looking at the data on the proper scale. Alternative techniques, methods, and instruments were employed to estimate the same quantity, necessitating a comparison of the strengths and limitations of the various alternatives. The quantity of data varies from one source to another and from one time period to another. The data contain both systematic errors for which adjustments must be made and random errors whose magnitude must be estimated. A high degree of autocorrelation exists in the data and this must be taken into account. The magnitude of error deviations and the conditions under which they are important require understanding and definition. Processing of the data requires the preparation of computation formulas for estimating the meteorological parameters in the most efficient manner. A knowledge of the accuracy with which the meteorological parameters are estimated will help in the design of future experiments.

2. ROLE OF BOOZ, ALLEN APPLIED RESEARCH INC.

Participation of Booz, Allen Applied Research Inc. (BAARINC) under the subject contract was initiated in October of 1969 following completion of the data-collection program. BAARINC was assigned the responsibility for developing a statistical methodology that would relate partially reduced BOMEX data to the end results required for budget computations in an optimum manner. This methodology requires the development of the rationale for estimation procedures and an understanding of the nature of error in the resulting estimate.

A substantial amount of data reduction is required to develop values of meteorological characters such as wind velocities, temperatures, and relative humidities at regular points along the path of a rawinsonde, dropsonde, or airplane path. BAARINC had the responsibility for developing a statistical plan for further processing of the data so as to arrive at estimates of the terms of the budget equation and to assess the precision of these estimates.

This final report covers the period from October 1969 through June 1970. An interim report covering the six-month period from October through March 1970 constitutes Appendix B of this report. The principal investigator for Booz, Allen Applied Research Inc. was Dr. Theodore W. Horner. Emphasis during the contract was placed on the design of a statistical plan, rather than the execution of such a plan. Such emphasis was partly due to the unavailability of data in a useful form throughout the study period. Samples of some data were available in preliminary form, but the extent of data available for inspection was strictly limited. The major portion of the data were unavailable because the analog data tapes had to be digitized, transformed to engineering units, edited, and reproduced in a form suitable for study.

Because of the unavailability of the data and the frontier nature of the BOMEX experiment, it was necessary to begin the development of statistical methodology on a trial and error basis. The methodology evolved as additional data became available and as the mathematics of the meteorological budget equations solidified.

Virtually none of this methodology had been developed in a recorded form before or during the BOMEX experiment, although Dr. Ben Davidson, the former scientific director, was probably well aware of such methodology. The untimely death of Dr. Davidson in

December preceding the BOMEX experiment created a substantial gap in analysis, reduction, and interpretation procedures. This gap has had to be filled by others.

After the end of the first six months of the subject contract, a substantial shift was made in the methodology under development. This shift occurred because of the recognition of the importance of the face average of a working box in the estimation process. The new concept introduces a much simpler and tighter methodology than that considered during the first six months. Since the latter does have value and is indicative of an alternative approach to the estimation of an interior average of a working box, it has been summarized as the interim report.

3. ESTIMATION OF BUDGET PARAMETERS

A central objective of the BOMEX experiment was to provide data useful for studying the flux of energy from the ocean to the atmosphere. This task is to be accomplished by keeping budgets on the passage into and out of the BOMEX box on the following properties:

- Mass
- Momentum-zonal
- Momentum-meridional
- Enthalpy
- Mechanical energy
- Total energy
- Latent heat budget.

The BOMEX box is located in an ocean area near the island of Barbados in the Caribbean. It is roughly 500 kilometers square and 500 millibars of pressure differential high. Extensive meteorological observations were collected around and within the box by means of ships, airplanes, satellites, and buoys.

4. ORGANIZATION OF THIS REPORT

Chapter II is a description of the end results required in the form of estimates of terms of the budget equations. Typical terms of the budget equations are analyzed with respect to calculative routines and the nature of the input data required to support these routines. The concepts of working boxes and interior and perimeter averages of working boxes are introduced.

The estimation of perimeter averages is discussed in Chapter III, first using rawinsonde data alone and then improving the estimates through the use of aircraft data.

Either the rawinsonde or the aircraft data may be subject to instrument bias; that is, the failure of the two systems to estimate the same quantity at the same space-time point. In Chapter IV, methods are described for investigating the nature of this bias, if any. In this chapter also, possible bias due to nonlinearity is discussed. The airplane data are useful for investigating this type of bias and for providing possible corrections to rawinsonde estimates of perimeter averages.

The estimation of interior averages is discussed in Chapter V and the methods of assessing the precision of estimates of the terms of the budget equations in Chapter VI. Chapter VII describes a plan for implementing the methodology described in the previous chapters and includes study conclusions and recommendations for future action.

The report contains two appendices. Appendix A covers a topic that is not directly related to the statistical plan, since it applies to a data problem that precedes the takeoff point for the study effort. This topic is that of developing a correction for rawinsonde humidity data due to heating of the hygristor sensing element. Appendix B is the interim report covering the first six months of the study effort.

II. ESTIMATION OF TERMS OF THE BUDGET EQUATIONS

BOMEX data are to be manipulated to estimate terms of the budget equations and to assess the precision of the estimates. Both point and confidence interval estimates are required. The budget equations are described in BOMEX Bulletin No. 6.*

As a basis for developing the required estimates, the BOMEX box will be partitioned into horizontal slices with a thickness of a 25 millibar pressure differential. These slices will be referred to as working boxes. Since the top of the BOMEX box is at 500-mb differential, there will be 20 such working boxes.

Two kinds of averages are required for each working box:

- Perimeter averages. This is an average of the values associated with all points around the perimeter of the working box.

- Interior averages. This is an average of values associated with all points constituting the working box.

For each type of average there are a number of characters (normal wind velocity, temperature, specific humidity) for which an average must be obtained. The character may be a product of two or more factors such as the product of specific humidity and the wind component normal to the side of the working box.

The ideal information for estimating the terms of the budget equations would be graphs of the true averages (perimeter and interior for each of the relevant characters) versus time, there being separate graphs for each working box. In place of the ideal information, the best possible graphs that can be developed from the BOMEX data will be employed. These approximation graphs are

* Environmental Science Services Administration, BOMEX Bulletin No. 6, Prepared by the BOMP Office, March 1970.

subject to various kinds of errors. The evaluation of these errors will afford the basis for evaluating the validity of the estimates of the terms of the BOMEX equations.

A perimeter graph will be defined as the graph of a perimeter average versus time. There will be one such graph for each working box or 20 such graphs for each character. Similarly, an interior graph is a graph of an interior average versus time. The time axis for these graphs will cover the entire period for which BOMEX data are available. Since the problems associated with constructing each kind of graph are different, the proposed methodology for each will be discussed separately. Rawinsonde, aircraft, and dropsonde data are all to be employed, in so far as possible, in the construction of an interior graph. Only rawinsonde and aircraft data are to be employed in the construction of the perimeter graphs.

Most of the computational features associated with using perimeter and interior graphs in estimating terms of the budget equations can be illustrated by examination of the following four typical terms of the budget equations.

1. FIRST TYPICAL TERM OF THE BUDGET EQUATIONS

The mass budget equation as given by Figure 12 of BOMEX Bulletin No. 6 is:

$$\frac{C}{A} \int_0^{P_T^*} [V_n] \frac{dp^*}{g} = - \frac{\overline{\omega}_T^*}{g} \qquad (2.1)$$

where

[ ] represents an average around the BOMEX box for the character inside the brackets

C = the circumference of the BOMEX box

A = the area of the BOMEX box

p^* = the difference in pressure at a given height and the pressure at sea level. p^* is $p_o - p$ where p is the pressure of the atmosphere at the height in question and p_o is the pressure at sea level

P_T^* = the value of p^* at the top of the box

$[V_n]$ = the average wind velocity around the perimeter of the box normal to the sides of the BOMEX box at the pressure differential height in question

g = the gravitational acceleration

$\overline{\omega}_T^*$ = dp^*/dt = the vertical velocity in the p^* system at the top of the box

The BOMEX data are to be manipulated so as to yield an estimate of the right-hand side of equation (2.1). Estimates of the right-hand side will be required at equally spaced times and as averages throughout specified time periods. The resulting estimates will be of meteorological interest in their own right and will also constitute input into other budget equations.

The basis for the estimation of the left-hand side is to approximate the integration over the pressure interval from sea level to the top of the BOMEX box by a summation over working boxes.

$$\frac{25\ C}{A\ g} \sum_{i=1}^{20} [V_{N_i}] = - \frac{\overline{\omega}_T^*}{g} \qquad (2.2)$$

The constant 25 in equation (2.2) is the pressure differential in millibars from the bottom to the top of a working box.

As input to the above equation, the perimeter graphs associated with the 20 working boxes are required. The character is the average velocity normal to the perimeter of the working i-th box; namely, V_{N_i}. Values of the averages from the graphs are required at specified times. In addition, the averages may be further averaged over specified time periods. The latter is defined by division of the area under average normal velocity curve by the length of the time period. Decisions are required as to the specified times and the time periods of interest.

2. SECOND TYPICAL TERM OF THE BUDGET EQUATIONS

A term of Figure 14 of BOMEX Bulletin No. 6 is:

$$\frac{\partial}{\partial t} \int_0^{P_T^*} \bar{H} \frac{dp^*}{g} \qquad (2.3)$$

where $\bar{H}$ is the average heat (enthalpy) in a working box located at a pressure level of p^*. The above term can be approximated as:

$$\frac{(25/g)}{(t_2 - t_1)} \sum_{i=1}^{20} \left[\bar{H}_{i2} - \bar{H}_{i1}\right] \qquad (2.4)$$

where

$\bar{H}_{i2}$ = average heat for ith working box at time t_2

$\bar{H}_{i1}$ = average heat for ith working box at time t_1

The $\bar{H}$ values required as input to (2.4) are taken from enthalpy interior graphs.

3. THIRD TYPICAL TERM OF THE BUDGET EQUATIONS

A second term from Figure 14 of BOMEX Bulletin No. 6 is:

$$\frac{C}{A} \int_0^{P_T^*} [H\ V_n] \frac{dp^*}{g} \qquad (2.5)$$

In this term, $[H\ V_n]$ is the perimeter average at a pressure height of p^* of the product of enthalpy (heat) and the normal wind velocity. This term can be treated in a manner similar to the treatment of the term on the left-hand side of the mass budget equation. In the present case the applicable character is $[H\ V_n]$ and the basis for the input required in (2. 5) are the perimeter graphs of the average product of H and V_n for each of the working boxes.

In employing the term under consideration in the equation of Figure 14 of BOMEX Bulletin No. 6, it is the average value of the term throughout a time interval (t_1, t_2) that is of interest. This means that the average of $[H\ V_n]$ throughout the time interval in question will be employed as input to (2. 5). The time interval (t_1, t_2) must be the same as that used to compute

$$\frac{\partial}{\partial t} \int_0^{P_T^*} \bar{\bar{H}} \frac{dp^*}{g} \qquad (2.6)$$

4. FOURTH TYPICAL TERM OF THE BUDGET EQUATIONS

A third term from Figure 14 of BOMEX Bulletin No. 6 is

$$-\int_0^{P_T^*} \left\{ \bar{\alpha}\, \bar{\omega}^* + \overline{\alpha''\, \omega^{*''}} \right\} \frac{dp^*}{g} \qquad (2.7)$$

where

$\bar{\alpha}$ = the average specific volume at a pressure height of p*

$\bar{\omega}''$ = the average vertical velocity

$\overline{\alpha''\, \omega^{*''}}$ = the covariance of α and ω^* on the p* surface

If the assumption is made that the covariance is zero, the term can be approximated as

$$- \int_{0}^{P_T^*} \bar{\alpha}\, \bar{\omega}^* \frac{dp^*}{g} \qquad (2.8)$$

The parameter $\bar{\alpha}$ is estimated from the relationship

$$\alpha = RT/p \qquad (2.9)$$

where

R = gas constant

T = temperature in degrees Kelvin

p = $- p^* + p_o$

Thus interior graphs for $\bar{\alpha}$ and $\bar{\omega}$ will be required for each working box.

III. DEVELOPMENT OF PERIMETER GRAPHS

The rawinsonde data for the development of the perimeter graphs are much more extensive than the aircraft data. The section that follows describes the procedure used for the construction of perimeter graphs based on rawinsonde data alone. The modification of this procedure to make use of aircraft data is also discussed.

1. EMPLOYMENT OF RAWINSONDE DATA

To provide a rationale for the proposed procedure, it is useful to first describe the structure of the available rawinsonde data and what should be done if each rawinsonde provided perfect information for the space-time continuum through which it passed. Second, the foregoing procedure will be modified to take into account the fact that the rawinsonde only estimates properties of the space-time continuum through which it passes.

(1) Perfect Rawinsonde Information

Rawinsondes were sent aloft at the corners and center of the BOMEX box at frequent intervals throughout the day and night. As one looks at a particular corner of the BOMEX box, there is a face of the box on one's left and a face on one's right. Each rawinsonde provided left face data and right face data. A particular rawinsonde cuts through all working boxes. Consider now a particular face, corner, and time interval. For the case in question, one can postulate a plot of a true value for the character versus the differential pressure p^*. The area under the plot divided by the pressure differential through the working box defines an average for the character in question. This average will be referred to as the rawinsonde right (or left) corner average.

It is postulated that the rawinsonde corner averages are known exactly and the question is now addressed as to how such information should be used to construct perimeter graphs. The proposed procedure is to make use of spike graphs. There will be one such spike graph for each combination of working box, left or right orientation, and character. For a given character there are thus 160 = (20 x 4 x 2) such graphs. A spike graph will relate the corner average to time as the abscissa. The height (ordinate) of the spike is associated with the corner average for the character. A spike will be located at the time (abscissa) that corresponds to the half-point time as the rawinsonde passes through the working box. The time of passage relative to the time between spikes is sufficiently small so that it can effectively be regarded as instantaneous. The assumption of perfect information is embodied in the assumption that the height (value) of each spike is known without error.

Each spike graph can be converted into a continuous graph by joining the tops of successive spikes. The eight continuous graphs provide the input data required for development of the perimeter graph for the working box. A perimeter average is formed by averaging the values taken from each of the eight continuous graphs for the same point in time. When a perimeter average, further averaged through a time interval, is desired, it can be obtained by averaging appropriate areas under the eight continuous curves.

The estimation of the perimeter average by the foregoing process is subject to error even if the corner averages are known exactly. To clarify the nature of this error, it is useful to introduce the concept of the <u>interval average</u>.

(2) Interval Averages

The rawinsonde corner (left or right) average is a special case of an interval average. In general if the value of a character is plotted over an interval, an interval average is the area under the plot divided by the length of the interval. In the case of the rawinsonde, the interval is a pressure interval passing through the working box. In the case of the airplane, the interval is the horizontal path of the airplane along the face of the working box. Theoretically, one can visualize a rawinsonde interval as existing at every point on a horizontal line drawn along the face of a working box. Conceptually, there is a plot of the rawinsonde interval average versus horizontal distance along the working box face. Actual rawinsonde data are found at only two points along this plot; namely, at the ends.

(3) Estimation Errors

Even with perfect rawinsonde information, the estimate of a perimeter average at a specified point of time is subject to error. There are two sources of error:

- Failure of the rawinsonde interval average to vary linearly over the face of the working box

- Failure of the interval averages to vary linearly with time for the time intervals between corner spikes.

Airplane data should prove useful in an examination of the validity of the first assumption. It is unlikely that the second assumption will hold exactly. It is our opinion that the loss in the precision of the estimate of a perimeter average resulting from any nonlinearily over the short time intervals involved should be small relative to other sources of variation.

At those space-time continuums where rawinsonde data are available, the interval average is not known exactly and must be estimated. Having chosen the best possible unbiased estimator for each spike in the spike graph, the conversion of the spike graph to a continuous graph can be improved by first smoothing the spike estimates.

This discussion of the estimation of a perimeter average through the use of rawinsonde data alone will be completed by discussion of estimators of spikes. The effects of errors in these estimates, along with other sources of errors, on the estimate of a perimeter average will be discussed in a separate chapter.

(4) Estimation of the Rawinsonde Corner Average

The optimum method of estimation of a rawinsonde corner average for a working box should take into account the structure and properties of the available data on which an estimate must be based. It is important also to be able to assess the precision of the resulting estimate.

The data available for estimating a corner average are values for the character in question at discrete points throughout the pressure interval of the working box. The discrete points are associated with five-second increments. Although the points are not equally spaced with respect to pressure differential, the discrepancies in spacing throughout a 25-mb interval are probably not important from the standpoint of estimation. At each data point throughout the corner interval, values are available for a set of characters such as temperature, specific humidity, and normal wind velocity.

It should be emphasized that the values of characters at specific points throughout the pressure interval that should be employed in estimating the corner average are the best values that can be made available. For example, it is known that relative humidities as recorded by the rawinsonde hygristor are incorrect because of heating of the hygristor sensing element due to solar radiation and the tubes and batteries of the unit in which the hygristor is located. Also there is a time lag that affects the observed relative humidity. In so far as possible, appropriate adjustments should be made to the observed values of characters to obtain new characters that will best reflect the meteorological quantity of interest. It is the new characters that should be employed in the estimation of interval averages.

The multitude of adjustments that must be made on old values to produce new values will tend to produce correlations among the error part of the "signal." The errors for the several characters are probably also correlated in some unknown manner. It is not likely that these correlations will be estimable or that their nature will be understood. Rather, it should be expected that such correlations will in fact be present, and that they may distort subsequent estimates and conclusions.

With respect to a particular character, three methods of estimation of the interval average are immediately obvious:

- Take an arithmetical average of the values at the discrete points
- Take a weighted average of the values of the discrete points, using as weights pressure differentials between points
- Fit a quadratic curve by least squares and use an estimate based on an appropriate function of the parameters of the fitting function.

Because spacing within the 25-mb pressure interval is almost uniform, there should be little difference between the first two methods. The second method should have slightly better precision. However, this additional precision is purchased at the expense of additional complexity. The principal objection to the first method is that it does not permit an estimate of the standard error of the estimate of the interval average. The usual way of computing the error variance as the ratio of the sample variance among the sample observations to the number of sample observations is not applicable because the sample variance among the observations includes the variation in the true signal throughout the interval and hence is too large.

In the third method, least squares procedures are employed to fit a quadratic model to the data. This method permits a calculation of the standard error of the estimate of the interval average. The calculated standard error should be largely free of upward bias due to systematic variation. The latter is that variation among the observations that is attributable to variation in the true value of the character along the pressure interval. The calculated standard error will have some small amount of upward bias due to the failure of the quadratic model to account completely for all systematic bias. On the other hand, the calculated standard error will have a downward bias due to the correlations in the errors of the observations. In general, if all of the observations in the pressure interval are too high or too low, this kind of error will not be reflected in the calculated standard error.

If there are n observations in the pressure interval, the standard error will be associated with n-3 degrees of freedom. The correction for the mean is associated with one lost degree of freedom of the three. The other two degrees of freedom are due to linear and quadratic variation of the character along the pressure interval. Thus, the result of removing most of the systematic effect, if any, is equivalent to losing two observations out of n.

In developing the standard error of the estimate of the interval average, we propose that an estimate applicable to a large class of such averages be developed by averaging (in the appropriate manner) the standard errors associated with many estimates. This will effectively increase the degrees of freedom to infinity.

Some of the mathematical details of the third method are sketched in the next two subsections.

(5) Quadratic Model

The following model is employed for the observations:

$$y_i = B_0 + B_1 x_i + B_2 x_i^2 + e_i \tag{3.1}$$

where B_0, B_1, and B_2 are constants and y_i is the observation along the pressure interval located at a coded distance x_i from the center of the pressure interval. The term e_i is an error deviation; that is, the difference between the value of the observation and the model. The e_i are assumed to have zero means and homogeneous variance. The normalized variable x_i is defined as

$$x_i = \frac{p_i^* - p_C^*}{\Delta p^*} \tag{3.2}$$

where

$$\begin{aligned} p_C^* &= (1/2)\,(p_U^* + p_L^*) \\ \Delta p^* &= (1/2)\,(p_U^* - p_L^*) \end{aligned} \tag{3.3}$$

and p_U^* and p_L^* are the upper and lower boundaries of the pressure interval.

Estimates of the constants B_0, B_1, and B_2 are determined by the usual least squares procedures. These are described in many textbooks. An excellent reference is Anderson and Bancroft*. The estimate of the interval average is

$$\hat{B}_0 + (1/3)\,\hat{B}_2 \tag{3.4}$$

where the $\wedge$'s (read hat) placed over a symbol indicate an estimate of the quantity underneath.

The variance of the estimate is

$$\sigma^2_{\hat{B}_0} + (2/3)\,\mathrm{Cov}\,(\hat{B}_0,\ \hat{B}_2) + (1/9)\,\sigma^2_{\hat{B}_2} \tag{3.5}$$

The variances and covariances of the estimates of the B parameters are derived by the usual procedures. Each can be expressed as a product of the variance from regression and a function of the pressure spacings of the points.

(6) Special Case of Points Equally Spaced

When the points of an interval can be regarded as effectively equally spaced, the formulas for estimating the properties of the interval are simplified. Thus, suppose that the first and last points of n points are located at distances of $1/n$ from the respective interval edges, and adjacent points are located at a distance from each other of $2/n$. The points would then be spaced along an interval running from -1 to 1 as $(1/n)-1$, $(3/n)-1$, ... $1-(3/n)$, $1-(1/n)$. The variable x in the equation

$$y = B_0 + B_1 x + B_2 x^2 + e$$

would take at these points the value given by the equation

$$x = [(2i - 1)/n] - 1 \tag{3.6}$$

* Anderson, R. L., and T. A. Bancroft, Statistical Theory in Research, McGraw-Hill, 1952.

The average of x^2 over the n points is

$$g = (1/n) \sum x^2 = 1/3 - 1/n^2 \qquad (3.7)$$

which is essentially 1/3 for $n \geq 5$.

In developing the properties of the interval, it is convenient to define two new variables, which sum to zero and which are uncorrelated, as

$$\begin{aligned} x_1 &= x \\ x_2 &= x^2 - g. \end{aligned} \qquad (3.8)$$

In terms of the new variables, the equation for y is

$$y = u + B_1 x_1 + B_2 x_2 + e \qquad (3.9)$$

where

$$u = B_0 + g\, B_2 \qquad (3.10)$$

In terms of the new model, the average value of y over all points is

$$\hat{u} = \bar{y} = 1/n \sum y_i = B_0 + g\, B_2 = B_0 + (1/3 - 1/n^2)\, B_2 \qquad (3.11)$$

Thus, the average of the observations over the interval is an estimate of the interval average, but it has a small bias.

Since x_1 and x_2 are uncorrelated and have zero means, the least squares equations for B_1 and B_2 are

$$\begin{aligned} \hat{B}_1 \sum x_1^2 &= \sum y x_1 \\ \hat{B}_2 \sum x_1^2 &= \sum y x_2 \end{aligned} \qquad (3.12)$$

It follows that the least squares estimates of B_1 and B_2 are

$$\hat{B}_1 = \frac{\sum y x_1}{\sum x_1^2}$$

$$\hat{B}_2 = \frac{\sum y x_2}{\sum x_2^2} \qquad (3.13)$$

The variances of the estimate are

$$\sigma_{\hat{u}}^2 = \sigma^2/n$$

$$\sigma_{\hat{B}_1}^2 = \sigma^2/\Sigma x_1^2 = \sigma^2/[n/3 - 1/n] = \sigma^2/ng \qquad (3.14)$$

$$\sigma_{\hat{B}_2}^2 = \sigma^2/\Sigma x_2^2$$

The covariances of these estimates are zero.

An unbiased estimate of the interval average can be computed as

$$\hat{u} + (1/n^2)\,\hat{B}_2 \qquad (3.15)$$

The variance of this estimate is

$$\sigma_{\hat{u}}^2 + (1/n^4)\,\sigma_{\hat{B}_2}^2 \qquad (3.16)$$

To estimate any of the above variances, an estimate of σ^2 is required. An estimate with n-3 degrees of freedom is

$$\sigma^2 = Z/(n-3) \qquad (3.17)$$

where

$$Z = \sum_i (y_i - \bar{y})^2 - \frac{\left(\sum_i yx_1\right)^2}{\sum_i x_1^2} - \frac{\left(\sum_i yx_2\right)^2}{\sum_i x_2^2} \qquad (3.18)$$

An improved estimate of σ^2 can be obtained by pooling the deviation sum of squares, the Z, over a number of intervals (j=1, 2, ...p) which are regarded as having somewhat similar error properties; for example, intervals for the same working box. A pooled estimate with $\sum_i (n_j - 3)$ degrees of freedom is

$$\sigma_P^2 = \frac{\sum_j Z_j}{\sum_j (n_j - 3)} \qquad (3.19)$$

If each interval has the same number of points such that $n_j = n$, then

$$\sigma_P^2 = \frac{\sum_j Z_j}{p\,(n - 3)} \qquad (3.20)$$

(7) Estimation of Derived Characters

Computations performed on the rawinsonde data give rise to observed point quantities such as specific humidity, temperature, and normal wind velocity at specific points along the rawinsonde path. These specific quantities are further employed to calculate derived point quantities such as the product of a normal wind velocity and specific humidity. Interval averages are also required for derived quantities.

The recommended procedure for forming interval averages of the derived quantities is to form the derived quantity at the available points along the interval. The derived quantity then becomes simply a new character that is operated on in the same manner as the observed characters. The interval averages obtained in this way are subject to an unknown bias. However, this bias will probably be small.

To show the nature of the bias, suppose characters y_i and z_i are observed at points i=1, 2, ...n. Error models for these characters are

$$y_i = g(x) + e_i$$

and (3.21)

$$z_i = h(x) + m_i$$

where g(x) and h(x) are the true value of the characters at point x on the pressure interval. The value of x proceeds from -1 at the bottom of the interval to zero at the middle and to 1 at the upper point of the interval. The terms e_i and m_i are error deviations that take up the slack between the observed and the true values. Ideally, we would like to estimate the average product of the true values over the interval; that is $\overline{g(x)\ h(x)}$. The proposed estimate in fact estimates

$$\overline{g(x)\ h(x)} + \mathrm{Cov}\ [g(x),\ m] + \mathrm{Cov}\ [h(x),\ e] + \mathrm{Cov}\ (e,\ m) \tag{3.22}$$

As long as y and z are different characters, it seems reasonable to assume that there is no correlation between the true value of one character and the error deviation of the second. Likewise there should be little correlation between the two errors. To the extent that the foregoing correlations are nonzero, the covariances in the above equation will be nonzero and the proposed estimate of the interval average will be biased.

In the special case where y and z are the same character, the estimate of the average value of y^2 over the interval will be subject to at least a bias of Cov $(e, e) = (\sigma^2)$; that is, the estimate will be biased upward by the error variance. In this case, the error variance must be estimated and subtracted from the interval average obtained by averaging the squares of the values.

2. EMPLOYMENT OF AIRCRAFT DATA

Aircraft data available for the estimation of perimeter averages is sparse compared to the rawinsonde data. Thus, aircraft data generally exists for only three working boxes whereas there are 20 such boxes. Even for the three working boxes, aircraft data are available for only a fraction of the time intervals for which rawinsonde data are available. Most of the information for the construction of perimeter graphs must be derived from the rawinsonde data.

The aircraft data can, however, be usefully employed in the following ways:

- To form an independent estimate of a perimeter average for comparison with a corresponding rawinsonde estimate
- To obtain an improved estimate of a perimeter average by proper combination of rawinsonde and aircraft data covering approximately the same time interval
- To develop corrections for rawinsonde perimeter averages.

The corrections will take into account nonlinear variation along the perimeter of the working box. The airplane data will provide the basis for such corrections.

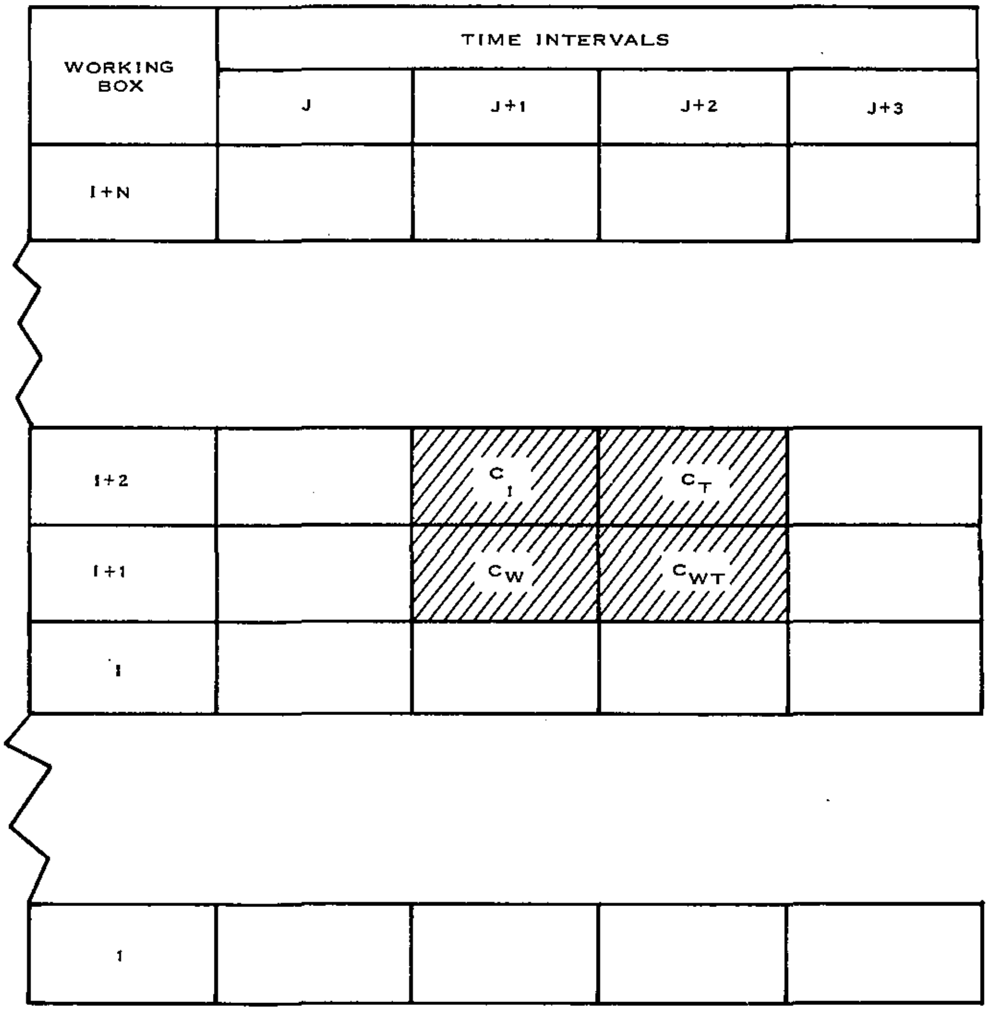
(1) Types of Corrections

The corrections that can potentially be applied to rawinsonde data can be classified according to the amount of extrapolation required in making the correction. Four levels of extrapolation are diagrammed in Figure 1. This figure applies to any selected side of the BOMEX box. The rows of the figure are associated with working boxes and the columns with time intervals.

The four corrections are listed below.

Correction	Description
C_1	Aircraft and rawinsonde data available for the same time interval and working face.

FIGURE 1
Levels of Extrapolation



NOTE: BOTH AIRCRAFT AND RAWINSONDE DATA ARE AVAILABLE IN CELL I+2, J+1. ONLY RAWINSONDE DATA ARE ASSUMED TO BE AVAILABLE IN THE OTHER SHADED CELLS.

Correction	Description
C_W	Aircraft and rawinsonde data available for the same time interval and for working boxes on the same face. However, extrapolation must be made to a different working box on the same face for the same time interval.
C_T	Aircraft data and rawinsonde data are available on the same working face at selected time intervals. Extrapolation is to a time interval not covered by the aircraft data.
C_{WT}	Aircraft and rawinsonde data are available for some working boxes on a specified face and for selected time periods. Extrapolation is to other working boxes on the same face and for other time intervals for which aircraft data are not available.

It is anticipated that the quality and structure of the data should be such that there will be little difficulty in making C_1 corrections. The likelihood that corrections C_W and C_T can be made, as of the writing of this report, appears excellent. On the other hand, difficulty is anticipated in forming C_{WT} corrections. Since it will be important for working box perimeter averages to be consistent from one working box to another and from one time period to another, it may not be worthwhile to attempt to apply C_1, C_W, and C_T corrections and not the C_{WT}. An answer to the question as to whether valid corrections can be formulated must await an appropriate analysis of the data along the lines indicated in subsequent chapters.

During the subject contract, considerable effort has been expended on methods of developing suitable corrections and employing such corrections. In our judgment it is very important to investigate the correction question. On the other hand, there is a large uncertainty as to whether corrections can be used to improve rawinsonde estimates of perimeter averages in a manner consistent for all working boxes and time intervals. Consequently our approach has been to develop a method for estimating perimeter averages that relies only on rawinsonde data. A secondary objective is to improve such estimates using aircraft and dropsonde data.

Methods for making corrections and comparisons between rawinsonde and aircraft data must take into account the nature of each type of data. The characteristics of the rawinsonde data have already been described. The nature of the aircraft data is described in the following subsection.

(2) Nature of the Aircraft Data

On various days and nights throughout the BOMEX experiment, aircraft flew at specified elevations along certain sides of the BOMEX box. A complicated pattern was followed--the airplane rising to an assigned elevation, flying along a side of the box at that elevation, moving to a different elevation, flying along the side at that elevation, and so forth. Generally, about one and one-half hours were required to fly a side. Several airplanes were used and they did not all have the same equipment. In many cases while one airplane flew at an assigned elevation on a given face, another airplane flew at the same elevation on the opposite face. The raw data collected by the airplanes were subject to various biases. Some of the information required for correction of biases was obtained from data on calibration squares flown at the corners of the box.

The starting point for the present analysis is a set of airplane data that reflects the results of all relevant adjustments. The data consist of sets of values of characters. Each set is associated with a different side interval. A side interval is the horizontal flight of an airplane along a face of the BOMEX box. Each side interval will be associated with a particular working box and period of time. Side intervals can thus be classified according to time, working box, and face. The set of data available for each side interval consists of values at a sequence of points along the interval. The distance between any two points is roughly associated with about five minutes of flight time. At each point there will be values for several characters.

(3) Face Averages

Since each side interval is identified with a particular face of a particular working box, it is useful to introduce the concept of the face average. The latter is an average of a character across a face of a working box. A working box perimeter average can be expressed as an average of the four face averages of the working box.

Although the ultimate object is to relate aircraft and rawinsonde perimeter averages, the object is best accomplished by relating estimates of rawinsonde and aircraft face averages. Thus it is easier and cleaner to synthesize a rawinsonde face average that covers essentially the same time frame as an airplace face average than it is to synthesize an airplane perimeter average that covers the same time frame as a rawinsonde perimeter average. Comparisons of aircraft and rawinsonde face averages, that relate to the same time frame, will determine whether there is a need to make corrections to the rawinsonde face averages and the nature of such corrections. Techniques for comparing rawinsonde and aircraft face averages and of developing corrections are discussed in a separate chapter.

The rawinsonde and aircraft data can be employed to develop the best face average graphs; that is, graphs of the face average versus time. Each graph will cover the entire BOMEX experiment for which relevant data are available. For each character there will be 80 such graphs (4 faces x 20 working boxes equals 80). The face average graphs should constitute the basis for the perimeter graphs.

An airplane interval average can be regarded as an estimator of the corresponding face average. The estimator suffers from the fact that variation of the character through the pressure interval is disregarded. The rawinsonde corner left and right corner averages can be averaged to form a second estimator of the face average. This latter estimator ignores horizontal variation along the face that is nonlinear. Comparison of the properties of the two types of estimators should permit a judgment as to the manner in which each estimator is deficient and consequently provide a basis for an improved estimator that properly combines rawinsonde and aircraft data.

The aircraft estimate of the face average is associated with a time duration of about 90 minutes. The rawinsonde corner averages will rarely occur in such a manner that there is a proper time correspondence between the rawinsonde face estimate and the aircraft face estimate. This problem can be surmounted by using a graph of the smoothed left corner average and the smoothed right corner average. The start and end times of the airplane flight can provide the times that are used to enter the graphs of the left and right corner averages.

(4) Estimation of Aircraft Side Averages

An aircraft side average is the average associated with an aircraft side interval. The techniques previously described for the estimation of a rawinsonde corner average are applicable to the estimation of an aircraft side average. The method of estimation through a quadratic model is recommended.

IV. METHODS OF DEVELOPING CORRECTIONS TO ESTIMATES OF RAWINSONDE FACE AVERAGES

Each aircraft interval is associated with a particular face of a particular working box, and each interval is identified with a time interval of approximately one and one-half hours duration. For each such face-time interval, aircraft data provide information as to variation of the character-in-question horizontally along the face. On the other hand, the rawinsonde data can be manipulated to provide information as to vertical variation across the same face. The two kinds of data can be combined to provide a fuller picture of variation across the face.

1. HORIZONTAL VARIATION ACROSS A FACE

To provide a basis for the rationale for the proposed full statistical model for characterizing variation over the face, we will discuss a model for variation horizontally across the face through the center. The model is

$$Y = C_0 + C_1 \; x + C_2^{\,2} \; x^2 \qquad (4.1)$$

where Y is the value of the character and the C's are constants. In the above equation, the variable x takes the value of 0 at the center and -1 and 1 at the ends of the horizontal line through the center. The model given in (4.1) should be distinguished from the statistical model

$$y = Y + e$$

where

y = an observation

Y = a function of coordinates and

e = an error deviation

The average value of Y across the horizontal line, which we will term the quadratic estimate, is

$$\overline{Y}_1 = (1/2) \int_{-1}^{1} Y\, dx = C_0 + (1/3)\, C_2 \tag{4.2}$$

In equation (4.2), the value of Y at the ends of the horizontal line are $C_0 - C_1 + C_2$ and $C_0 + C_1 + C_2$. The average value of Y along a straight line joining the two end points is $C_0 + C_2$. This average value will be termed the linear estimate and will be denoted as $\overline{Y}_2$, thus

$$\overline{Y}_2 = C_0 + C_2 \tag{4.3}$$

The linear estimate can be regarded as an approximation to the quadratic estimate. The relationship between the two estimates is

$$\overline{Y}_1 = \overline{Y}_2 - (2/3)\, C_2 \tag{4.4}$$

The left and right rawinsonde corner points estimate $C_0 - C_1 + C_2$ and $C_0 + C_1 + C_2$ respectively. The rawinsonde estimates (or the aircraft estimates) may be subject to unknown biases. Although efforts will have been made at this stage to have removed such biases, it will be necessary to make provisions for the existence of any biases that may remain in the model. Letting Y_{RL} and Y_{RR} denote the left and right rawinsonde corner points and Δ's represent biases,

$$\begin{aligned} C_0 - C_1 + C_2 &= Y_{RL} + \Delta_L \\ C_0 + C_1 + C_2 &= Y_{RR} + \Delta_R \end{aligned} \tag{4.5}$$

Thus

$$C_0 + C_2 = (1/2)\,(Y_{RL} + Y_{RR}) + (1/2)\,(\Delta_L + \Delta_R) \tag{4.6}$$

The rawinsonde estimate of the average value of Y across the horizontal line is

$$\bar{Y}_3 = (1/2)\,(Y_{RL} + Y_{RR}) \qquad (4.7)$$

Thus the relationship between the quadratic estimator and the rawinsonde estimator is

$$\bar{Y}_1 = \bar{Y}_3 + (1/2)\,(\Delta_L + \Delta_R) - 2/3\; C_2 \qquad (4.8)$$

Thus, two types of bias may be present in the rawinsonde estimator:

- Instrument bias. This bias occurs if the rawinsonde and the aircraft estimate different quantities at the same space time point, namely an end of the horizontal line
- Nonlinear bias. This is the failure of the character to vary in a linear manner along the horizontal line.

In studying the effects of the two types of bias, we are not only concerned with whether statistical evidence exists for such a bias in terms of significance tests, but whether the bias has practical importance. The magnitudes of the ratios $(1/2)\; \bar{Y}_3)\; (\Delta_L + \Delta_R)$ and $(2/3\; \bar{Y}_3)\; C_2$ throw light on the practical importance. Hopefully these ratios will be found to be sufficiently small so that $\bar{Y}_3$ can be regarded as a reasonable approximation to $\bar{Y}_1$.

(1) Instrument Bias

The two types of bias can be studied separately. To study the instrument bias, we consider the statistic (see (4.4))

$$\bar{Y}_2 = \bar{Y}_1 + (2/3)\; C_2 = C_0 + C_2 \qquad (4.9)$$

which can be estimated entirely from aircraft data. This statistic is to be compared to $\bar{Y}_3$, which can be estimated entirely from rawinsonde data. The estimate of the difference $\bar{Y}_2 - \bar{Y}_3$ is an estimate of $(1/2)(\Delta_L + \Delta_R)$. Techniques for determining the properties of $(1/2)(\Delta_L + \Delta_R)$ will be described in section 4 of this chapter.

(2) Nonlinear Bias

The nonlinear bias can be studied by estimating C_2 for each aircraft interval using aircraft data. The estimates are then to be examined for systematic changes with height and other relevant variables using analysis of variance, regression, and other statistical techniques. Ideally, it is desirable to develop a model that would predict C_2 as a function of variables whose values are readily known. This prediction equation would then form the basis for making corrections to the rawinsonde estimates.

2. MODEL FOR VARIATION OVER A FACE

Variation over the face of a working box may be nonlinear not only in a horizontal direction, but in a vertical direction. In this section we formally develop a model that makes full use of the available information on the horizontal and vertical variation over a face. In the next section we show that the analysis of the full model leads to essentially the same end result as the analysis based on the horizontal line through the center of the face.

The full model, however, is useful if an estimate of the variance over the face is desired. This variance is defined as

$$\sigma_F^2 = (1/4)\int_{-1}^{-1}\int_{-1}^{-1} Y^2 dx_1\, dx_2 - \left\{(1/4)\int_{-1}^{-1}\int_{-1}^{-1} Y\, dx_1\, dx_2\right\}^2 \qquad (4.10)$$

A sufficient model for characterizing the vertical variation is probably the quadratic model

$$Y = B_0 + B_1 x + B_2 x^2 \qquad (4.11)$$

where Y is the value of the character at a vertical distance x from the center. In the preceding model, the variable x is assumed to have the value of zero at the center and -1 and 1 at the lower and upper edges of the face.

Estimates of the values of the B's will be available for the left and right edges of the face from the rawinsonde data. No data are available as to variation in the vertical direction of the face at other intermediate horizontal points along the face. It appears reasonable, however, to approximate the vertical variation associated with any point along the horizontal center line of the face by linear interpolation on constants reflecting vertical variation at the edges of the face.

Thus suppose x_1 represents the horizontal distance across the face with x_1 taking the value of 0 at the center and -1 and 1 at the left and right edges. The coefficients of vertical variation at horizontal distance x_1 can be approximated as

$$B_0 = (1/2)\,[B_{L0}\,(1 - x_1) + B_{R0}\,(1 + x_1)] \tag{4.12}$$

$$B_1 = (1/2)\,[B_{L1}\,(1 - x_1) + B_{R1}\,(1 + x_1)] \tag{4.13}$$

$$B_2 = (1/2)\,[B_{L2}\,(1 - x_1) + B_{R2}\,(1 + x_1)] \tag{4.14}$$

where L and R identify constants of the model for quadratic variations at the left and right edges of the face. For the special cases of the left edge, the center, and the right edge, the above constants of vertical variation reduce as shown below.

Case	x_1	B_0	B_1	B_2
Left edge	-1	B_{L0}	B_{L1}	B_{L2}
Center	0	$(1/2)(B_{L0}+B_{R0})$	$(1/2)(B_{L1}+B_{R1})$	$(1/2)(B_{L2}+B_{R2})$
Right edge	1	B_{R0}	B_{R1}	B_{R2}

To develop an equation that combines horizontal and vertical variation over a face, equation (4.12) is rewritten as

$$B_0 = (1/2)(B_{L0} + B_{R0}) + (1/2)(B_{R0} - B_{L0})\, x_1 \qquad (4.15)$$

We now identify C_0 as $(1/2)(B_{L0} + B_{R0})$ and C_1 as $(1/2)(B_{R0} - B_{L0})$. By incorporating a term for quadratic variation in the horizontal direction and letting x_3 denote* vertical distance, the full model for variation over the face can be written

$$\begin{aligned} Y = {} & C_0 + C_1 x_1 + C_2 x_1^2 \\ & + (x_3/2)\,[B_{L1}(1 - x_1) + B_{R1}(1 + x_1)] \qquad (4.16) \\ & + (x_3^2/2)\,[B_{L2}(1 - x_1) + B_{R2}(1 + x_1)] \end{aligned}$$

The vertical average as a function of x_1 is

$$\bar{Y} = C_0 + C_1 x_1 + C_2 x_1^2 + (1/6)\,[B_{L2}(1 - x_1) + B_{R2}(1 + x_1)]$$

3. APPLICATION OF THE FULL MODEL FOR VARIATION OVER A FACE

The average value of the character over the face is

$$\bar{\bar{Y}} = (1/4) \int_{-1}^{1} \int_{-1}^{1} Y \, dx_1 \, dx_3 \qquad (4.17)$$

Using equation (4.16) this average reduces to

$$\bar{\bar{Y}}_1 = C_0 + (1/3)\, C_2 + (1/3) \left[\frac{B_{L2} + B_{R2}}{2} \right] \qquad (4.18)$$

* The x_2 symbol will be reserved for horizontal variation perpendicular to the x_1 axis.

A model for variation over the face based on a model for the average information contained only in the rawinsonde data is

$$
\begin{aligned}
Y = {} & (1/2)\,[B_{L0}\,(1 - x_1) + B_{R0}\,(1 + x_1)] \\
& + (x_3/2)\,[B_{L1}\,(1 - x_1) + B_{R1}\,(1 + x_1)] \\
& + (x_3^2/2)\,[B_{L2}\,(1 - x_1) + B_{R2}\,(1 + x_2)]
\end{aligned}
\tag{4.19}
$$

The average for this model over the face is

$$
\bar{\bar{Y}}_3 = (1/2)\,(B_{L0} + B_{R0}) + (1/3)\left[\frac{B_{L2} + B_{R2}}{2}\right] \tag{4.20}
$$

This average is also the same as the average of the corner averages. The left and right corner averages are

$$
\begin{aligned}
&(1/2)\int_{-1}^{1} (B_{L0} + B_{L1}x_3 + B_{L2}x_3^2)\,dx_3 = B_{L0} + (1/3)\,B_{L2} = \bar{Y}_{RL} \\
&(1/2)\int_{-1}^{1} (B_{R0} + B_{R1}x_3 + B_{R2}x_3^2)\,dx_3 = B_{R0} + (1/3)\,B_{R2} = \bar{Y}_{RR}
\end{aligned}
\tag{4.21}
$$

where the bars over Y_{RL} and Y_{RR} distinguish the corner averages from the values at ($x_1 = -1$, $x_3 = 0$) and ($x_1 = 1$, $x_3 = 0$). Values at these latter points were denoted in (4.5) as Y_{RL} and Y_{RR}. Under the full model, the rawinsonde estimate is the average of the averages; that is,

$$
\bar{\bar{Y}}_3 = (1/2)\,(\bar{Y}_{RL} \text{ and } \bar{Y}_{RR}) = (1/2)\,(B_{L0} + B_{R0}) + (1/3)\left[\frac{B_{L2} + B_{R2}}{2}\right] \tag{4.22}
$$

The above equation for $\bar{Y}_3$, which is based on the full model, replaces (4.7) that applied only to the model for variation along the horizontal center line of the face.

The rawinsonde and the aircraft data have two points in common as follows:

Point	x_1	x_3	Full Model	Rawinsonde Model
Center of left edge	-1	0	$C_0 - C_1 + C_2$	B_{L0}
Center of right edge	1	0	$C_0 + C_1 + C_2$	B_{R0}

At these two points, the aircraft and the rawinsonde may be estimating somewhat different quantities because of bias in one or both of two types of data. The values given by the two models at the two points relate in the following manner

$$\begin{aligned} C_0 - C_1 + C_2 &= B_{L0} + \Delta_L \\ C_0 + C_1 + C_2 &= B_{R0} + \Delta_R \end{aligned} \qquad (4.23)$$

where the Δ's represent bias terms. Thus

$$C_0 + C_2 = (1/2)\,(B_{L0} + B_{R0}) + (1/2)\,(\Delta_L + \Delta_R) \qquad (4.24)$$

and

$$C_0 + (1/3)\,C_2 = (1/2)\,(B_{L0} + B_{R0}) + (1/2)\,(\Delta_L + \Delta_R) - (2/3)\,C_2\,. \qquad (4.25)$$

Equation (4.18) can now be rewritten as

$$\bar{\bar{Y}}_1 = (1/2)\,(B_{L0} + B_{R0}) + (1/3)\left(\frac{B_{L2} + B_{R2}}{2}\right) + (1/2)\,(\Delta_L + \Delta_R) - (2/3)\,C_2\,.$$

By referring to (4.20) for the representation of $\bar{\bar{Y}}_3$, it follows that the average for the full model relates to the model for the average for the rawinsonde model in the following way

$$\bar{\bar{Y}}_1 = \bar{\bar{Y}}_3 + (1/2)\,(\Delta_L + \Delta_R) - (2/3)\,C_2\,.$$

This is the same result as (4.8) of section 1 of this chapter.

4. EXAMINATION OF INSTRUMENT BIAS

As defined by equation (4. 3) and estimated from airplane data, $\bar{\bar{Y}}_2$, and, as defined by equation (4. 7) and estimated from rawinsonde data, $\bar{\bar{Y}}_3$ presumably estimate the same quantity. It should be possible to calculate a large number of pairs of estimated values of $\bar{\bar{Y}}_2$ and $\bar{\bar{Y}}_3$. To relate the two estimates and to determine the nature of any difference between them, the following model can be employed

$$\begin{aligned} u_i &= \bar{\bar{Y}}_{3i} \\ v_i &= \bar{\bar{Y}}_{2i} \\ y_i &= u_i + e_i \\ z_i &= v_i + h_i \end{aligned} \tag{4. 26}$$

where

y_i = the value of the rawinsonde estimate for the i-th pair

z_i = the value of the aircraft estimate for the i-th pair

u_i = the quantity estimated by the rawinsonde estimate

v_i = the quantity estimated by the aircraft estimate

e_i = a random error in the rawinsonde estimate distributed with mean zero and variance $\sigma_R{}^2$

h_i = a random error in the aircraft estimate distributed with mean zero and variance $\sigma_A{}^2$.

The pairs are to be examined so as to find out the following:

- Are the u_i and v_i the same, and, if not, do they differ in any systematic way?
- Are the error variances of e_i and h_i the same, and, if not, what are their relative magnitudes?

The first question relates to the bias $(1/2)(\Delta_L + \Delta_R)$, whereas the second question examines the relative precision of the two types of estimates and the proper method of combination.

(1) Relationship Between the u_i and the v_i

To gain insight into the first question, u_i is plotted versus z. The resulting plot, apart from sampling fluctuations, will be a straight line passing through the origin with a 45 degree slope if u_i tends to be the same as v_i. It is a simple matter to test if the foregoing assumptions are valid. Ordinary least squares regression techniques cannot be employed, however, since an important assumption will not hold; namely, that only one of the variables be subject to error. Wald-Bartlett[*] [†] procedures can be employed for testing the hypotheses that the line has a 45 degree slope and passes through the origin.

It is to be hoped that the data will support the following equality between u_i and v_i; that is, $u_i = v_i$. It is perhaps more likely that the observed relationship will take the form

$$u_i = A + B\, v_i \tag{4.27}$$

where A is an intercept constant and B is a slope constant. Thus, the relationship may be linear in form but with A other than zero and B other than 1. If the data supports a relationship more complicated than (4.27), it is unlikely that corrections for rawinsonde data can be formulated. If the data supports (4.27), then estimates of A and B should be available which will permit either y or z to be adjusted so that they will at least estimate the same quantity.

* Wald, A., "The Fitting of Straight Lines If Both Variables Are Subject to Error," Ann. Math. Stat., 11:284ff, 1940.

† Bartlett, M. S., "Fitting a Straight Line When Both Variables are Subject to Error," Biometrics, 5:207-212, 1949.

(2) <u>Comparison of Error Variances Resulting from Covariance Analysis</u>

With respect to the second question, the sample variances and covariances can be computed for the two types of estimates y and z. These will provide estimates of the true variances and covariances. The latter will have the following structure.

$$\begin{aligned} \sigma_y^2 &= \sigma_u^2 + \sigma_R^2 \\ \sigma_z^2 &= \sigma_v^2 + \sigma_A^2 \\ \operatorname{Cov}(y, z) &= \operatorname{Cov}(u, v) \end{aligned} \tag{4.28}$$

Now if u_i is the same quantity as v_i, the following relationship will hold:

$$\sigma_u^2 = \sigma_v^2 = \operatorname{Cov}(u, v) \tag{4.29}$$

Hence

$$\begin{aligned} \sigma_y^2 &= \sigma_u^2 + \sigma_R^2 \\ \sigma_z^2 &= \sigma_u^2 + \sigma_A^2 \\ \operatorname{Cov}(y, z) &= \sigma_u^2 \end{aligned} \tag{4.30}$$

The preceding three equations can be solved for σ_R^2, σ_A^2, and σ_u^2.

The component σ_u^2 represents that portion of the variation in the observations that is attributable to variation among the true values that are being estimated. If σ_u^2 is found to be small relative to σ_y^2 or σ_z^2 then the following interpretations should be considered:

- A nonlinear relationship between the y and z
- Small variation among the true values relative to the noise in the estimates of those true values
- Consistency of the true values among the examined cases.

The final possibility can be further examined by computing the ratio $(\sigma_u)/\bar{y}$. The foregoing constitutes a coefficient of variation. A small coefficient of variation indicates that the variation among the true values from case to case is small relative to the mean value among the cases.

The preceding types of analyses should be applied to various groups of pairs so as to verify the consistency of the relationship of u_i to v_i from one group to another. For example, does the same relationship hold from one working box to another? At this point in time, we might anticipate that either A or B of (4. 27) would vary in a systematic manner with height.

(3) Variance Comparisons

The analysis of the foregoing section provides an estimate of the variance of y and an estimate of the variance of z. These two variances were denoted as σ_R^2 and σ_A^2 respectively. By methods to be described in a subsequent chapter, it is possible to estimate the variances of y and z by an independent method. The second method is based on estimated variances of data points from the quadratic model over intervals. If members of the pair of variance estimates of y (or z) are similar in magnitude, then this will increase our confidence that a rawinsonde interpolated estimate estimates the same quantity as the airplane estimate. If the variances are substantially different, then it is probably the case that

$$u_i = v_i + \eta_i \tag{4. 31}$$

where η_i is a random variable with unknown properties.

(4) Combining of Estimates

Both y and z contain information on $C_0 + C_2$. If the two estimates are found to be unbiased relative to each other, it makes sense to combine the two estimates for those face-periods where both types of data are present. The combining should be accomplished by weighting each estimate inversely according to the error variance. The combined estimate for $C_0 + C_2$ will be represented as x_1. The equation for x_1 is

$$x_1 = \frac{w_y y + w_z z}{w_y + w_z} \tag{4.32}$$

where $w_y = 1/\sigma_R^2$ and $w_z = 1/\sigma_A^2$

The estimate of the face average is then

$$x_2 = x_1 - (2/3)\, C_2 \tag{4.33}$$

V. DEVELOPMENT OF INTERIOR GRAPHS

A methodology for the development of interior graphs is described in this chapter. An interior graph is a graph of the interior average of the character-in-question over the x_1, x_2, x_3 coordinates of the working box. The data foundation for preparation of the interior graphs must rely principally on the rawinsonde data since aircraft and dropsonde data are sparse in comparison. The latter two types of data, at best, will permit refinements of the interior graphs prepared from rawinsonde data. The principal refinement will be that of taking into account curvilinear variation of the character throughout the working box.

The methodology for using rawinsonde data alone in the preparation of interior graphs will be described first. This will be followed by a discussion of a model for the variation throughout the interior of the working box. Refinements that employ aircraft data and refinements that employ dropsonde data will then be discussed in turn.

1. EMPLOYMENT OF RAWINSONDE DATA

In obtaining an interior average it is convenient to first average in the vertical direction and then to average horizontally. The vertical average should be obtained as an average of the quadratic curve over the vertical interval $x_3 = -1$ to 1. The quadratic curve is:

$$Y = B_0 + B_1 x_3 + B_2 x_3^2$$

and the vertical average is

$$\bar{Y} = (1/2) \int_{-1}^{1} Y \, dx_3$$

$$= B_0 + (1/3) B_2$$

Estimates of the vertical average for various combinations of the x_1 and x_2 coordinates are available from the dropsonde and rawinsonde data. In particular for the rawinsonde data, estimates of the vertical average are available for the x_1, x_2 corner combinations below.

x_1	x_2
1	1
1	-1
-1	1
-1	-1

The vertical averages at the corners of the working box were referred to in Chapter III as rawinsonde corner averages. Each average can be graphed versus time and hence corner graphs prepared. The corner graphs are spike graphs, and all of the discussion of section 1 of Chapter III is applicable to such graphs and need not be repeated here. For each character there are four such graphs for each working box, rather than eight, as was the case in the estimation of perimeter averages, since a distinction between the right and left corner averages is unnecessary in the estimation of interior averages.

The corner graphs are to be transformed into continuous graphs by joining spikes or smoothing as appropriate. An interior graph is prepared by reading values associated with the same time instant from the four corner graphs and averaging the values read. In actual operation, it will be the computer equivalent of the foregoing process that will be employed.

The method of preparing interior graphs described above employs all of the information contained in the rawinsonde data. This information is described by the following statistical model for variation of the vertical average over the x_1, x_2, plane. This model is

$$\bar{Y} = D_0 + D_1 x_1 + D_2 x_2 + D_{12} x_1 x_2 \tag{5.1}$$

where the D's are constants. The model provides for a constant that characterizes the general level, a linear effect over x_1, a linear effect over x_2, and a linear by linear effect. An average of the Y values over the x_1, x_2 plane is

$$\bar{\bar{\bar{Y}}} = (1/4) \int_{-1}^{1} \int_{-1}^{1} \bar{Y}\, dx_1\, dx_2 = D_0 \qquad (5.2)$$

The model is deficient in that it does not take into account possible effects of curvilinear variation of $\overline{Y}$ over the x_1, x_2 plane. Dropsonde and aircraft data provide hope of correcting this deficiency by supplying a correction to the interior average estimated from rawinsonde data alone. Although it is important to research the basis for corrections, it is debatable, as of the time of preparation of this report, as to whether useful corrections can be formulated in view of the characteristics of the data. Characteristics, with respect to the usefulness of corrections, will only be known at a subsequent stage of data reduction.

2. MODEL FOR VARIATION OVER THE x_1, x_2 PLANE

A more general model than (5.1) for variation of $\overline{Y}$ over the x_1, x_2 plane is the two-dimensional quadratic model

$$\bar{Y} = G_0 + G_1\, x_1 + G_2\, x_2 + G_{11}\, x_1^2 + G_{12}\, x_1\, x_2 + G_{22}\, x_2^2 \qquad (5.3)$$

where the G's are constants and x's are the coordinates of points in the x_1, x_2 plane. The coordinates have the value of zero at the center of the BOMEX box and -1 and -1 on the edges of the box. The two-dimensional quadratic model is at best only an approximation for the true model for the variation of $\overline{Y}$ over the x_1, x_2 plane. However, the BOMEX data do not appear to be sufficient to support a model more complex than (5.3). In our judgment, this latter model will extract essentially all of the relevant information contained in the dropsonde and aircraft data for improving the rawinsonde estimate of the interior mean.

Using the two dimensional quadratic model, the interior mean of the working box is

$$\bar{\bar{\bar{Y}}} = (1/4) \int_{-1}^{1} \int_{-1}^{1} \bar{Y}\, dx_1\, dx_2 = G_0 + (1/3)\,(G_{11} + G_{22})\,. \qquad (5.4)$$

(1) Four Corner Average

The interior mean as given by (5.4) can be contrasted to a mean calculated as the average of the four corner values. Such a contrast is useful in developing the informational content of the dropsonde data. The four corner values as developed from (5.3) are as follows.

x_1	x_2	
1	1	$G_0 + G_1 + G_2 + G_{11} + G_{12} + G_{22}$
1	-1	$G_0 + G_1 - G_2 + G_{11} - G_{12} + G_{22}$
-1	1	$G_0 - G_1 + G_2 + G_{11} - G_{12} + G_{22}$
1	1	$G_0 - G_1 - G_2 + G_{11} + G_{12} + G_{22}$

The average of the four corner values is

$$\bar{\bar{\bar{Y}}}_2 = G_0 + G_{11} + G_{22} \qquad (5.5)$$

Thus the interior average relates to the <u>four corner average</u> through the equation

$$\bar{\bar{\bar{Y}}}_1 = \bar{\bar{\bar{Y}}}_2 - (2/3)\,(G_{11} + G_{22}) \qquad (5.6)$$

The rawinsonde data will provide an estimate of $\bar{\bar{\bar{Y}}}_2$. This estimate is an average of the four rawinsonde corner averages. The dropsonde data would also provide some information on $\bar{\bar{\bar{Y}}}_2$, except it is known that the dropsonde data are subject to a bias. There appears to be little hope at this time of estimating the magnitude of this bias. The dropsonde data should, however, prove useful in estimating $(2/3)\ (G_{11} + G_{22})$.

(2) Perimeter Average

The perimeter average based on (5.3) can be contrasted to the interior mean to provide insight into possible use of aircraft data in estimating the interior mean. Using (5.3), the four face averages are derived as follows.

$$\begin{aligned} &G_0 - G_2 + G_{22} + G_{11}/3 \\ &G_0 + G_2 + G_{22} + G_{11}/3 \\ &G_0 - G_1 + G_{11} + G_{22}/3 \\ &G_0 + G_1 + G_{11} + G_{22}/3 \end{aligned} \tag{5.7}$$

The perimeter average of the working box is thus

$$\bar{\bar{\bar{Y}}}_5 = G_0 + (2/3)\ (G_{11} + G_{22})\ . \tag{5.8}$$

Using (5.4) it follows that the interior mean relates to the perimeter average by the equation

$$\bar{\bar{\bar{Y}}}_1 = \bar{\bar{\bar{Y}}}_5 - (1/3)\ (G_{11} + G_{22})\ . \tag{5.9}$$

The aircraft data should provide estimates of both $\bar{\bar{\bar{Y}}}_5$ and $(1/3)\ (G_{11} + G_{22})$.

In passing it should be noted that usually it will be the case that either aircraft data or dropsonde data will be present for a working box-time interval, but not both. When both are present, each can be employed to estimate $G_{11} + G_{22}$. The two estimates can be combined inversely according to their error variances. In those cases where neither estimate is present, an extrapolation must be made from those cases where one or the other or both are present. This extrapolation must be such that consistency is maintained with respect to adjoining boxes and time intervals.

3. EMPLOYMENT OF AIRCRAFT DATA

The aircraft data will also provide information on a quantity estimated by the rawinsonde corner average through

$$\overline{\overline{\overline{Y}}}_6 = \overline{\overline{\overline{Y}}}_5 + (1/3)\,(G_{11} + G_{22}) \tag{5.10}$$

which is $G_0 + G_{11} + G_{22}$, which is the same quantity, apart from bias, as $\overline{\overline{\overline{Y}}}_2$ in (5.5). The pair of estimates $\overline{\overline{\overline{Y}}}_2$ and $\overline{\overline{\overline{Y}}}_6$ should be available for a number of cases, which are classified in various ways. The relationships between $\overline{\overline{\overline{Y}}}_2$ and $\overline{\overline{\overline{Y}}}_6$ can be studied in the manner set out in section 4 of Chapter IV. If bias is not found, then the rawinsonde estimate of $\overline{\overline{\overline{Y}}}_2$ and the aircraft estimate of $\overline{\overline{\overline{Y}}}_6$ can be combined by weighing the estimates inversely according to their error variances.

In the absence of bias and when rawinsonde, dropsonde, and aircraft data are all present for a working box-time interval, perhaps the best way to combine all of the data is to first obtain the best combined estimate of $G_{11} + G_{22}$ using dropsonde and aircraft data. We will represent this combined estimate as $\widehat{G_{11} + G_{22}}$. One-third of the combined estimate is then added to the estimate of the aircraft perimeter average to obtain a $\hat{\overline{\overline{\overline{Y}}}}_7$. Thus

$$\hat{\overline{\overline{\overline{Y}}}}_7 = \hat{\overline{\overline{\overline{Y}}}}_5 + (1/3)\,\widehat{G_{11} + G_{22}} \tag{5.11}$$

The estimate $\hat{\bar{\bar{Y}}}_7$ is then combined with the rawinsonde corner estimate $\hat{\bar{\bar{Y}}}_2$ by weighing inversely according to variances to form a $\hat{\bar{\bar{Y}}}_8$. The best combined estimate of the interior mean is then

$$\hat{\bar{\bar{Y}}}_9 = \hat{\bar{\bar{Y}}}_8 - (2/3)\ \widehat{(G_{11} + G_{22})} \tag{5.12}$$

The estimate $\bar{\bar{Y}}_5$ is an estimate of the perimeter average. This was discussed in Chapter III. The aircraft estimate of $G_{11} + G_{22}$ is an average of the four quadratic coefficients that are associated with the four aircraft intervals along the perimeter of the working box.

4. EMPLOYMENT OF DROPSONDE DATA

Dropsonde observations potentially can provide information on curvilinear variation of the character-in-question over the x_1, x_2, x_3 coordinates of the working box. The usefulness of the dropsonde observations for this purpose is limited by the nature of the dropsonde data. The discussion of the use of dropsonde data follows the sequence:

- Nature of the dropsonde observations
- Transformation or coordinates
- Diagonal face averages.

(1) Nature of the Dropsonde Observations

Dropsondes were dropped along the diagonals of the BOMEX box. In a particular airplane run eight dropsondes were dropped, four along each diagonal. The four dropsondes on each diagonal were distributed two on one side of the center and two on the other. The first and last dropsonde along the diagonal were each about two-thirds of the distance from the center to their respective corners. A typical pattern for an airplane flight is as follows.

Diagonal	Time	Elapsed Time
First	0204 Start	
		59 minutes
	0303 End	
		76 minutes
Second	0419 Start	
		53 minutes
	0512 End	
		3 hours + 8 minutes

Thus, completion of each diagonal required approximately an hour. Completion of both diagonals required approximately three hours. This time is sufficiently small, relative to the times over which major changes in interior box averages occur, that little loss of information should accrue if effects due to time changes during completion of the two diagonals are disregarded. Estimates of the effect on the interior box average due to curvilinearity should be assigned an instantaneous time that is in the middle of the airplane flight.

As a dropsonde passes through a working box, it provides a picture of the variation of the character-in-question throughout a vertical interval of the working box. A quadratic function can be fitted to the data of the vertical interval and an estimate of the mean for the vertical interval obtained in the form

$$\bar{Y} = B_0 + (1/3)\, B_2$$

For each aircraft flight there will be eight such values for each working box. It is known that the eight values were subject to a constant, but unknown bias, that varied randomly from one aircraft flight to another. Thus, the only useful information in the eight vertical means has to do with the failure of these eight means to vary in a linear manner over the x_1, x_2 plane.

(2) Transformation of Coordinates

In working with dropsonde data, it is convenient to employ a different coordinate system in which the axes of the new system correspond to the diagonals of the x_1, x_2 plane. In the new system the origin remains at the center of the BOMEX box. The new coordinates, which will be denoted as x_1' and x_2', will range from $-\sqrt{2}$ to $\sqrt{2}$ at the ends of the diagonals. The new coordinates relate to the old coordinates of (5. 3) by the equations

$$x_1 = (1/\sqrt{2})(x_1' - x_2')$$

and (5. 13)

$$x_2 = (1/\sqrt{2})(x_1' + x_2') .$$

In terms of the new coordinates, the two-dimensional quadratic model is

$$\overline{Y} = G_0' + G_1' x_1' + G_2' x_2' + G_{11}' (x_1')^2 + G_{12}' x_1' x_2' + G_{22}'(x_2')^2 . \quad (5.14)$$

This model, as would be expected, has the same structure as (5.3). The new constants relate to the old constants through the equations

$$\begin{aligned} G_{11}' &= (1/2)(G_{11} + G_{22} + G_{12}) \\ G_{22}' &= (1/2)(G_{11} + G_{22} - G_{12}) \\ G_{12}' &= G_{22} - G_{11} . \end{aligned} \quad (5.15)$$

Thus, it is evident that

$$G_{11}' + G_{22}' = G_{11} + G_{22} . \quad (5.16)$$

The coefficients G_{11}' and G_{22}' are coefficients of quadratic variation along each of the diagonals of the working box. Thus, data along one diagonal can be employed to estimate G_{11} and data along the other diagonal can be employed to estimate G_{22}. The estimate of $(2/3)(G_{11} + G_{22})$ is obtained as the estimate of $(2/3)(G_{11}' + G_{22}')$. The latter estimate is simply $(2/3)(\widehat{G}_{11}' + \widehat{G}_{22}')$.

To obtain the estimates of G_{11}' and G_{22}', the four dropsonde averages along the diagonal can be employed. A simple quadratic model is fitted to these four points. The estimate of the quadratic coefficient in this latter model constitutes the required estimate of G_{11} or G_{22}, as the case may be.

(3) Examination of Instrumentation Bias in the Dropsonde Observations

Instrumentation bias is expected in the dropsonde observations. It is unlikely that very much can be done to adjust the dropsonde observations for the effect of this bias. Even so it is important to study the nature and magnitude of this bias. This can be done in a manner similar to that employed in the examination of instrumentation bias in the estimation of perimeter face averages. A diagonal face average will, however, replace the perimeter face average.

A diagonal face can be visualized as that face formed by a vertical plane cutting a working box through diagonal corners. The vertical average along this face can be expressed in the form

$$\bar{Y} = C_0 + C_1 x'' + C_2 (x'')^2 \tag{5.17}$$

where the C's are constants and x'' now varies from zero at the center to -1 and 1 at the ends of the diagonal face. The average over the face is

$$\bar{Y}_1 = C_0 + (1/3) C_2 \tag{5.18}$$

This average relates to an average based on rawinsonde diagonal corner averages by the equation

$$\bar{Y}_1 = \bar{Y}_3 + 1/2 (\Delta_L + \Delta_R) - (2/3) C_2 \tag{5.19}$$

where the terms in this latter equation have the same definitions as the terms in sections 2 and 3 of Chapter IV. However, here the corners diagonally opposite each other are employed instead of the corners of a perimeter face. In estimating $\bar{Y}_3$ it is important to choose properly the times at which the corner rawinsonde are read.

A dropsonde estimate free of curvature bias is

$$\overline{Y}_2 = \overline{Y}_1 + (2/3)\ C_2 = C_0 + C_2 \qquad (5.20)$$

which estimates the same quantity, apart from instrumentation bias as the rawinsonde estimate $\overline{Y}_3$. The relationships between $\overline{Y}_2$ and $\overline{Y}_4$ can be studied along the lines indicated in section 4 of Chapter IV.

The x'' value in this subsection was permitted to vary from -1 to 1 so as to parallel the discussion of section 1 of Chapter IV. In subsection 2 of this chapter, an x' varied along the diagonal from $-\sqrt{2}$ to $\sqrt{2}$. The relationship between x' and x'' is

$$x' = \sqrt{2}\, x''$$

It follows that

$$C_2\ (x'')^2 = B'\ (x')^2 = B'\ 2\ (x'')^2$$

where the B' is either the B_{11}' or the B_{22}' and the x' is either the x_1 or the x_2' of section 2 of this chapter. Thus

$$B' = (1/2)\ C_2 \qquad (5.21)$$

The computation of $(2/3)\ (B_{11} + B_{22})$ required as input to (5.6) is most easily accomplished by computing

$$\widehat{(2/3)\ (B_{11} + B_{22})} = (1/3)\ [C_2 \text{ for one diagonal} + C_2 \text{ for the other diagonal}] \qquad (5.22)$$

VI. COMPUTATION OF STANDARD ERRORS AND CONFIDENCE LIMITS

It is important not only to estimate the terms of the budget equations, but also to assess the precision of the estimates in terms of standard errors and confidence limits. The details of any assessment procedure are intimately related to the precise way in which the terms of the budget equations are estimated. Since detailed specifications will not be formulated until sometime after submission of this report, it is the object of this chapter to indicate principles for assessment.

Confidence intervals will first be discussed. This will be followed by a section on linear functions, which constitute the basic principle on which variance derivations are based. A more comprehensive discussion of linear functions will be found in Anderson and Bancroft, which has been previously cited. The application of linear functions in the present context is illustrated by three examples, two of which are in the section on linear functions. These two examples demonstrate the computation of the variance of the estimate of the budget term given by (2.2) when the estimate is based entirely on rawinsonde data. The third example, which is in a separate section, derives the variance of an estimator of the interior mean of a working box that combines aircraft, dropsonde, and rawinsonde data, specifically the estimate given by (4.33).

1. CONSTRUCTION OF CONFIDENCE INTERVALS

The precision of an estimate of a term of the budget equation is best indicated in terms of confidence limits. Upper and lower $1 - \alpha$ confidence limits can be constructed from the equation below.

Lower limit

$$\hat{B} - T_{\alpha} s_{\hat{B}}$$

$$\tag{6.1}$$

Upper limit

$$\hat{B} + T_{\alpha} s_{\hat{B}}$$

$\hat{B}$ is an estimate of the term of the budget equation in question and $s_{\hat{B}}$ is the estimated standard error of $\hat{B}$. The parameter T_α is defined as that value which is exceeded, with probability $1 - \alpha$, by the absolute value of a normal deviate. Symbolically,

$$\mathcal{P}(|T| \geq T_\alpha) = 1 - \alpha \qquad (6.2)$$

where $\mathcal{P}$ indicates probability and T is a normal deviate.

The confidence limits are interpreted as providing the end points of an interval that will cover the true value of the budget term being estimated unless a rare event has occurred. The probability of this rare event is $1 - \alpha$. The confidence limits have been formulated in terms of the normal deviate, instead of Student's t, since it is anticipated that $s_{\hat{B}}$ will generally be estimated with sufficient degrees of freedom so that those degrees of freedom can be regarded as infinite. Most statisticians regard thirty degrees of freedom sufficient for this purpose. We will insure that the degrees of freedom in the present context will be larger than thirty by pooling homogeneous error sums of squares in the manner of (3.19). The T_α values associated with different α values are set out below.

T	α
1.000	0.68
1.645	0.10
1.960	0.05
2.576	0.01

The standard error $s_{\hat{B}}$ can be computed as the square root of the estimated variance of $\hat{B}$. Subsequent discussion will focus on estimating variances.

2. LINEAR FUNCTIONS

The estimates of the terms of the budget equations are expressible as linear functions of interval averages and curvature estimates; that is,

$$y = \sum_i a_i z_i \qquad (i = 1, 2 \ldots p) \tag{6.3}$$

where y represents the linear function (budget term estimate), p is the number of terms in the linear function, the a_i are constants and the z_i are random variables. The latter will either be an interval average or a curvature estimate for an interval. The variance of the linear function is

$$\sigma_{\hat{B}}^2 = \sum_i a_i \; \sigma_{z_i}^2 + \sum_{i} \sum_{\substack{i' \\ i' \neq i}} a_i \, a_i \; \text{Cov} \, (z_i, \; z_{i'}) \tag{6.4}$$

where

$$\sigma_{z_i}^2 = \overline{(z - \bar{z})^2} \tag{6.5}$$

and

$$\text{Cov} \, (z_i, z_i') = \overline{(z_i - \bar{z}_i) \, (z_i' - \bar{z}_i)} \tag{6.6}$$

Some special cases of the linear function should help to clarify the properties of (6.4).

Case	Variance	Comment
$y = a\, z$	$\sigma_y^2 = a^2\, \sigma_z^2$	The variance of the product of a constant and a random variable is equal to the product of the square of the constant and the variance of the random variable.
$y = z_1 \pm z_2$	$\sigma_y^2 = \sigma_{z_1}^2 + \sigma_{z_2}^2$	The variance of the sum or difference of two uncorrelated random variables is the sum of the variances.

$$y = a_1 z_1 - a_2 z_2 \quad \sigma_y^2 = a_1^2\, \sigma_{z_1}^2 - 2\, a_1\, a_2\, \mathrm{Cov}\,(z_1,\ z_2) + a_2^2\, \sigma_{z_2}^2$$

The variance of the linear function is estimated by substituting estimates of the variances and covariances of the random variables.

(1) Variance of an Estimate of a Budget Term

As an example of application of (6.4), consider the variance of the budget term given by (2.2). The variance is

$$\left(\frac{25/g}{t_2 - t_1}\right)^2 \sum_{i=1}^{20} \left(\sigma_{\hat{\bar{H}}_{i2}}^2 + \sigma_{\hat{\bar{H}}_{i1}}^2\right) \tag{6.7}$$

where $\sigma_{\hat{\bar{H}}_{i2}}^2$ and $\sigma_{\hat{\bar{H}}_{i1}}^2$ are the variances of the interior averages $\hat{\bar{H}}_{i2}$ and $\hat{\bar{H}}_{i1}$. In this application, two assumptions have been made:

- Estimates of interior averages are uncorrelated from one working box to another
- Estimates of interior averages are uncorrelated from one time instant to another.

The first assumption appears reasonable if the assumption is made that each working box average is an unbiased estimate of the true working box average. The error deviation of the estimate from the true value will then be a function of the error deviations of character values from quadratic models. The second assumption should be valid if the time instants are sufficiently far apart so that the estimates are derived from data on different rawinsondes.

(2) <u>Variance of an Estimate of an Interior Mean</u>

The working box interior average $\overline{H}_i$ may be estimated as the average of four corner averages. Thus

$$\hat{\overline{H}}_i = (1/4) \sum_j y_j \qquad (6.8)$$

where y_j is the estimate of the j-th (i = 1, 2, 3, 4) corner interval average. The variance of $\hat{\overline{H}}_i$ is

$$\sigma_{\hat{\overline{H}}_i}^2 = (1/16) \sum \sigma_{y_j}^2 \qquad (6.9)$$

No covariances were included in the right-hand side of the above formula because the corner interval averages should be uncorrelated from one corner to another with respect to their error components. The error components should be uncorrelated because they are associated with different rawinsondes.

To estimate a corner average for the precise instant of time required may require a linear interpolation between two corner averages at the two successive times that inclose the time instant in question. Thus an estimate of y_j may be calculated as

$$y_i = a_1 y_{j1} + a_2 y_{j2} \tag{6.10}$$

where

$$a_1 = 1 - \frac{t' - t_1}{t_2 - t_1} \tag{6.11}$$

and

$$a_2 = \frac{t' - t_1}{t_2 - t_1} \tag{6.12}$$

and where y_{j1} and Y_{j2} are estimated averages, calculated from data on different rawinsondes, at times t_1 and t_2 respectively. Generally $t_2 - t_1$ will be of the order of 90 minutes. An estimate of the corner average is required at t'. The variance of y is

$$\sigma_{y_j}^2 = a_1^2 \sigma_{y_{j1}}^2 + a_2^2 \sigma_{y_{j2}}^2 \tag{6.13}$$

where $\sigma_{y_{j1}}^2$ and $\sigma_{y_{j2}}^2$ are variances of the averages used to calculate y_j. No covariance term was included in the above formula because y_{j1} and y_{j2} are calculated from data on different rawinsondes and hence should be uncorrelated.

The formula for the variance of y_j can be simplified for practical use. First, the variances of y_{j1} and Y_{j2} should be similar, and hence it is reasonable to replace the individual variance by a pooled variance. This enables $\sigma_{y_j}^2$ to be written as

$$\sigma_{y_j}^2 = \sigma^2 \left(a_1^2 + a_2^2\right). \tag{6.14}$$

Second, there will be innumerable cases where y_j estimates will be required. Across these innumerable cases, t' can lie at any point between t_1 and t_2. Thus it appears reasonable to replace $a_1^2 + a_2^2$ by an average value developed as t' is allowed to vary at random between t_1 and t_2. The average value is

$$a_1^2 + a_2^2 = 2/3. \tag{6.15}$$

This average value appears reasonable in view of the values of $a_1^2 + a_2^2$ calculated for the special values of t' listed below.

t'	$a_1^2 + a_1^2$
t_1	1
$(1/2)(t_1 + t_2)$	1/2
t_2	1

The recommended formula for the variance of y_j is

$$\sigma_{y_j}^2 = (2/3)\,\sigma_j^2 \tag{6.16}$$

where σ_j^2 is the variance of the average associated with the j-th corner. If the further assumption is made that the σ_j^2 are homogeneous, that is $\sigma_j^2 = \sigma^2$, the variance of the working box interior average reduces to

$$\sigma_{\hat{H}}^2 = (1/16)\,(4)\,(2/3)\,\sigma^2 = (1/6)\,\sigma^2 \tag{6.17}$$

where σ^2, as employed here, denotes the variance of $\hat{B}_0 + (1/3)\hat{B}_2$. In the case of equally spaced points, σ^2 is

$$\sigma^2 = \sigma_P^2 \left[\frac{1}{n} + \left(\frac{1}{n^4}\right)\left(\frac{1}{S\,x_2^2}\right)\right] \tag{6.18}$$

where σ_P^2 is given by (3.20).

3. VARIANCE OF A COMBINED ESTIMATOR OF AN INTERIOR MEAN

An estimator was developed, along with its rationale, in section 3 of Chapter V, which employed rawinsonde, dropsonde, and aircraft data.

The estimator requires the input information listed below:

z_1 = an estimate of the rawinsonde corner average

z_2 = an estimate of the aircraft perimeter average

z_3 = an estimate of $G_{11} + G_{22}$ based on the aircraft data

z_4 = an estimate of $G_{11} + G_{22}$ based on dropsonde data

$\hat{\sigma}_1^2$ = estimated variance of z_1

$\hat{\sigma}_2^2$ = estimated variance of z_2

$\hat{\sigma}_3^2$ = estimated variance of z_3

$\hat{\sigma}_4^2$ = estimated variance of z_4

It will be assumed that all of the z's apply to the same instant of time. Each z may be calculated as a function of more basic quantities. Thus, z_1 may be calculated as the average of four corner interval averages. Each of the latter may be a linear interpolation estimate based on interval averages whose times bracket the time instant in question. The variance of z_1 is also a function of more basic variances. The z's will be assumed to be uncorrelated. This appears to be reasonable because data from different sources are associated with the different z's with the exception of z_2 and z_3. However, z_2 and z_3 should also be uncorrelated if the method of Chapter III, section 1.6 is employed as a basis for the estimation process.

In symbolically expressing the combined estimator, it is convenient to define the z's below.

$$
\begin{aligned}
z_5 &= W_{11} z_3 + W_{12} z_4 \qquad W_{11} + W_{12} = 1 \\
z_6 &= z_2 + z_5/3 \qquad\qquad\qquad\qquad\qquad\qquad (6.19) \\
z_7 &= W_{21} z_1 + W_{22} z_6 \qquad W_{21} + W_{22} = 1 \\
z_8 &= z_7 - (2/3)\, z_5
\end{aligned}
$$

The combined estimator is z_8. It is formed by first calculating z_5 by combining the dropsonde and aircraft estimates of $G_{11} + G_{22}$. The method of combination is to weight inversely according to the variances; that is,

$$W_{11} = \frac{1/\sigma_3^2}{1/\sigma_3 + 1/\sigma_4^2} = \frac{\sigma_4^2}{\sigma_3^2 + \sigma_4^2} \tag{6.20}$$

$$W_{12} = \frac{1/\sigma_4^2}{1/\sigma_3^2 + 1/\sigma_4^2} = \frac{\sigma_3^2}{\sigma_3^2 + \sigma_4^2} = 1 - W_{11} \tag{6.21}$$

Using (6.4), the variance of z_5 is derived as

$$\sigma_5^2 = \frac{\sigma_3^2 \quad \sigma_4^2}{\sigma_3^2 + \sigma_4^2} \tag{6.22}$$

One third of z_5 is then added to the airplane perimeter average to obtain a z_6, which estimates the same quantity as the rawinsonde average z_1. The variance of z_6 is

$$\sigma_6^2 = \sigma_2^2 + (1/9)\,\sigma_5^2 \tag{6.23}$$

The z_7 statistic is formed by weighting z_6 and z_1 inversely according to their error variances. The weights are

$$\begin{aligned} W_{21} &= \frac{1/\sigma_1^2}{1/\sigma_1^2 + 1/\sigma_6^2} = \frac{\sigma_6^2}{\sigma_1^2 + \sigma_6^2} \\ &= \frac{\sigma_2^2 + (1/9)\left(\dfrac{\sigma_3^2 \quad \sigma_4^2}{\sigma_3^2 + \sigma_4^2}\right)}{\sigma_1^2 + \sigma_2^2 + (1/9)\left(\dfrac{\sigma_3^2 \quad \sigma_4^2}{\sigma_3^2 + \sigma_4^2}\right)} \end{aligned} \tag{6.24}$$

$$W_{22} = \frac{1/\sigma_6^2}{1/\sigma_1^2 + 1/\sigma_6^2} = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_6^2} = 1 - W_{21}$$

Finally, the combined estimate is formed by subtracting (2/3) z_5 from z_7.

If z_8 is expanded in terms of its basic components, we have

$$z_8 = W_{21} z_1 + W_{22} [z_2 + (1/3) (W_{11} z_3 + W_{12} z_4)] - (2/3) (W_{11} z_3 + W_{12} z_4)$$

or

$$z_8 = W_{21} z_1 + W_{22} z_2 + z_3 (W_{11}/3) (W_{22} - 2) + z_4 (W_{12}/3) (W_{22} - 2) \tag{6.25}$$

Again using (6.4), the variance of z_8 is derived as

$$\sigma_8^2 = W_{21}^2 \sigma_1^2 + W_{22}^2 \sigma_2^2 + (W_{11}/3)^2 (W_{22} - 2)^2 \sigma_3^2 + (W_{12}/3)^2 (W_{22} - 2)^2 \sigma_4^2 \tag{6.26}$$

If a z_1, z_2, z_3, or z_4 is missing, then the σ^2 associated with that z becomes infinite and the weight associated with the component becomes zero. This forces the other weight of the pair to 1. The value of σ_8^2 is estimated by substituting in estimates of σ_1^2, σ_2^2, σ_3^3, and σ_4^2.

VII. EXECUTION OF THE STATISTICAL PLAN

The estimation of the various terms of the budget equations will require considerable computation and manipulation of data files. In this chapter a procedure is sketched for performing the required work in an orderly manner. The procedure will provide the basis for the development of a computer program for accomplishing the calculations. This chapter also contains conclusions concerning the statistical aspects of budget term estimation and recommendations for insuring that such estimation is accomplished in an optimum manner.

1. TASKS PRELIMINARY TO ESTIMATION OF BUDGET TERMS

Estimation of the terms of the budget equations is based on the statistics of intervals (average, curvature, variances,...). By and large, the budget estimates are linear functions of estimates of interval statistics.

A number of tasks must be accomplished and a number of decisions must be made before the main computational process for estimation of budget terms can be initiated. These preliminary tasks are as follows.

(1) Computational Analysis of Budget Terms

Each term of the budget equation should be assigned a code number for subsequent reference and each term should be analyzed from a computational point of view; that is, a specific computational formula should be formulated that will replace the theoretical formula for the budget term.

The computational formulas will generally take one of the two forms below.

Form A (time interval quantity)

$$B = \sum_{i=1}^{20} C_i \, (y_{it'} - y_{it})$$

where

B = the budget term

C_i = a constant associated with the ith box working box

y_{it} = an interior average at time t. t' is a time subsequent to t.

Form B

$$B = \sum_{i=1}^{20} C_i \, y_{i,t \text{ to } t'}$$

where

B = the budget term

C_i = a constant associated with the ith working box

$y_{i,\, t \text{ to } t'}$ = an average value of a perimeter average throughout the time interval t to t'.

The budget terms are classified into time instant quantities and time interval quantities. A time instant quantity is one that requires estimates at specific instants of time. Generally, there are two instants in time and it is the difference between values at the two time instants that is of interest. A time interval quantity is the average value of a budget term throughout a specified time interval.

Specific numerical values need to be defined for the C_i constants. Computational formulas in terms of interval statistics must be further specified for the y_i averages. Associated with each budget term will be one or more characters whose values will enter into computation of the term. These characters must be listed, since they form the basis for a list of characters on which interval statistics must be computed. The computational analysis of the budget terms will provide a signature for the budget term in terms of a combination of codes:

- Code for the budget term
- Codes for forms of budget terms
- Codes for characters
- Codes for constants
- Code for position in order of computation.

(2) Develop Interval Statistic Reference System

Budget term estimates will be formed by combining the statistics of intervals. Procedures must be developed for selection of the appropriate statistics in application of computational formulas. The first step in providing a selection procedure is the development of an interval statistic reference system in the form of codes. The system should have two aspects:

- Identification of the interval
- Identification of the statistic within the interval.

Factors to be considered with respect to interval identification are listed in Table 1. The statistics within each interval can be further classified according to character. Thus a character code is required. This code should be further supplemented by a code differentiating between observed characters and derived characters. Each statistic that is associated with a character must be identified. Table 2 is a listing of the statistics that are required for each character. This list should be reviewed and updated as research continues in Project BOMAP.

(3) Calculate and Interpret Interval Statistics

Lists of the characters must be prepared, both observed and derived, for which interval statistics are required. These interval statistics must then be calculated, identified, assembled in an orderly arrangement, and made readily accessible through a medium such as a tape or cards. The assembled and readily accessible set of interval statistics will constitute the set of building blocks on which budget terms estimates will be based. Ideally the tape should be ordered according to

Table 1
Factors for Interval Identification

Type of data

- Rawinsonde
- Aircraft
- Dropsonde

Particular application

- Which rawinsonde of the rawinsondes
- Which aircraft flight along which face
- Which dropsonde of the dropsondes

Working box

Position

Rawinsonde	Aircraft	Dropsonde
Corner	Face	Diagonal
Left or right		Location within diagonal

Midpoint time of interval

Table 2
Statistics Required for Each Character for Each Interval

Estimates	Variances
Interval average	Variance of the interval average
$\hat{B}_0$	$\hat{\sigma}^2_{\hat{B}_0}$
$\hat{B}_1$	$\hat{\sigma}^2_{\hat{B}_1}$
$\hat{B}_2$	$\hat{\sigma}^2_{\hat{B}_2}$
	Variance from regression

Other	Covariances*
n	Matrix of sums of squares and cross products
$\sum_i x_1^2$	Inverse matrix
$\sum_i x_2^2$	
Start and end points of interval	

* Only required if the points are not equally spaced.

- Character
- Working box within character
- Type of data (rawinsonde, dropsonde, aircraft)
- Position within type of data and working box
- Time within position
- Statistics within times and position.

The building block tape should first be exercised by having the computer print out spike graphs for visual inspection and interpretations. There should be a separate spike graph for each combination of interval statistic, position, and working box. These graphs should bring to light any difficulties with the interval statistics. Subsequent data reduction should only be continued after any observed difficulties have been resolved. The standard errors of the more important interval statistics should be employed in the interpretation of the spike graphs.

(4) Development of Corrections

The spike graphs, particularly those of curvature statistics, should be useful in the development of corrections to rawinsonde interior and perimeter averages. Methods of formulating corrections were described in Chapter IV of this report. Ideally, these methods will provide a new set of correction graphs, which will be similar in form to the spike graphs.

(5) Choice of Time Instants and Time Intervals

Decisions are required for the time instants at which, and the time intervals for which, budget terms should be computed. Such decisions should be possible after examination of the spike graphs. The estimated standard errors of the interval statistics will influence greatly such decisions.

2. DESIGN OF COMPUTER PROGRAM FOR ESTIMATION OF BUDGET TERMS

A computer program for estimating terms of the budget equation will first be described based on rawinsonde data alone.

(1) Rawinsonde Data Alone

It will be assumed that rawinsonde interval averages are arranged according to character (i), working box within character (j), position within working box (k), and according to time within position. For data employed in the calculation of perimeter averages, there will be eight positions associated with left and right orientations in each of four corners. For data employed in the calculation of interior averages, there will be four positions associated with the four corners. The computations are to be completed by varying j within i and then moving to the next value of i.

Within each ij combination, the following steps are to be completed.

Step 1

For each position, develop a new set of interval averages. The new averages are the best estimates of interval averages at specified instants of time or the best estimates of averages of interval averages over specified periods of time. If the former, then the computer must locate two interval averages whose midpoint times bracket the time instant in question. The new interval average can then be formed by linear interpolation or through application of some appropriate weighting procedure. To compute the average of the interval averages, the computer locates those interval averages within the time period in question and forms an appropriate average of them. The new averages will be represented by $\bar{y}_{ipjk}$ where p references the new averages according to time order.

Step 2

Average the new interval averages across positions; that is, form

$$\bar{y}_{ipj} = (1/g) \sum_{k} \bar{y}_{ipjk}$$

where g is 4 for interior characters and 8 for perimeter characters.

Step 3

Multiply $\bar{y}_{ipj}$ by an appropriate constant so that a meaningful divergence (or interior) statistic ($\bar{z}$) for the working box is formed. Thus a z_{ipj} is

$$\bar{z}_{ipj} = c\, \bar{y}_{ipj}$$

where c is a constant.

Step 4

Calculate

$$w_{ipp'j} \doteq z_{ip'j} - z_{ipj}$$

if it is the difference between the z's at specified times that is required.

Step 5

Accumulate the $\bar{z}_{ipj}$, or the $\bar{w}_{ip'pj}$, as the case may be, across the working boxes to form budget estimates ($\hat{B}$) for the i^{th} character and p^{th} time slot (or p, p' time interval). Thus,

$$\hat{B}_{ip} = \sum_j \bar{z}_{ipj}$$

$$\hat{B}_{ip'p} = \sum_j \bar{w}_{ip'pj}$$

(2) Use of Aircraft and Dropsonde Data

If aircraft and dropsonde data are found to be useful in improving the foregoing procedure, the only modification required is a modification of step 2 by adding a correction factor to $(1/g) \sum y_{ipik}$

3. CONCLUSIONS

The following conclusions were drawn with respect to the statistical treatment of BOMEX data during the contract period.

(1) The structure of the BOMEX data is such that estimates of budget terms are possible.

(2) Standard errors and confidence limits can be constructed for estimates of the budget terms.

(3) Estimates of budget terms can be made using only rawinsonde data.

(4) Dropsonde and aircraft data can improve the rawinsonde estimates of the budget terms by adjusting for the effects of curvilinearity.

(5) The key concepts of the estimation process are

- Working boxes
- Interior averages of working boxes
- Perimeter averages of working boxes
- Face averages of working boxes
- Interval averages.

(6) Techniques are available for examination of bias due to instrument error and nonlinearity.

(7) A major step in budget computations is the preparation of a building block tape. This tape will contain coded descriptions of intervals and statistics associated with each interval.

(8) The building block has a logical order with respect to characters, working boxes, data types, positions, and times.

(9) A computer program can be designed which will perform the calculations required for budget estimates in a logical order using the building block tape as input data.

4. RECOMMENDATIONS

To enhance the statistical benefits of the BOMEX data the following recommendations are made.

(1) Execution of preliminary tasks described in section 1 of this chapter should be initiated.

(2) Programming to implement the foregoing concepts should be initiated as rapidly as possible, particularly with respect to the following:

- Calculation and identification of interval statistics
- Design of the building block tape
- Computer preparation of spike graphs using the building block tape
- Preparation of a program for estimating terms of the budget equations using the building block tape.

(3) A standard error study should be initiated. Preliminary estimates of the order of magnitude of standard errors

are now possible with the available data using the techniques described in this report.

(4) At an appropriate time a study should be conducted as to the inferences to future BOMEX type experiments that can be drawn as conclusions from the statistical analysis of the BOMEX data.

REFERENCES

1. Environmental Science Services Administration, BOMEX Bulletin No. 6, Prepared by the BOMAP Office, March 1970.

2. Anderson, R. L., and T. A. Bancroft, Statistical Theory in Research, McGraw-Hill, 1952.

3. Wald, A., "The Fitting of Straight Lines If Both Variables Are Subject to Error," Ann. Math. Stat., 11:284ff, 1940.

4. Bartlett, M. S., "Fitting a Straight Line When Both Variables Are Subject to Error," Biometrics, 5:207-212, 1949.

APPENDIX A

A METHOD FOR ESTIMATING THE TEMPERATURE OF THE HYGRISTOR SENSING ELEMENT

Procedures have been formulated to estimate the relative humidity in the immediate neighborhood of the sensing elements of a rawinsonde hygristor. Unfortunately the sensing element does not correspond to the relative humidity of the surrounding atmosphere because of the higher temperature of the air immediately adjacent to the sensing element. The higher temperature results from the interplay of long-wave radiation in the duct between the hygristor and the duct walls. The long-wave radiation is due to the heating of the radiosonde. The latter is heated from batteries and tubes and incident solar radiation. The incident solar radiation is highly dependent upon the thickness and density of intervening clouds that prevent solar radiation from reaching the radiosonde. The actual temperature of the atmosphere is available from a thermistor also attached to the radiosonde. This temperature is considered to be good after appropriate corrections for lag are made. The hygristor relative humidities will possess errors due to heating even after application of lag corrections. The end problem is to estimate the specific humidity of the atmosphere.

The approach described here is based on comparing the properties of a distribution of observations derived from airplane data with properties of a similar distribution based on rawinsonde data. Differences in the properties permit an estimate of a correction to be applied to the rawinsonde thermistor temperature so that the corrected temperature will be a good estimate of the temperature of the hygristor sensor element.

The basis of the proposed procedure hinges heavily on the equation that relates vapor pressure, relative humidity and temperature. The next section describes this relationship. The following section develops the mathematical basis of the proposed correction and the final section covers various comments about the method.

1. VAPOR PRESSURE

A basic assumption in the derivation that follows is that the moisture vapor pressure in the immediate vicinity of the hygristor sensing element of a rawinsonde is unaffected by any excess of temperature of the air in the immediate vicinity over and above that of the surrounding atmosphere. This means that the relative humidity recorded by the hygristor is given by the formula

$$R = \left\{ e/\left[(6.11)\,(10^{at_H/(b+t_H)})\right]\right\} \times 100 \qquad (A.1)$$

where

R = relative humidity recorded by the hygristor

e = vapor pressure of air in immediate vicinity of the hygristor sensing element

t_H = the temperature of the air in the immediate vicinity of the hygristor sensing element expressed in degrees centigrade

a = 7.5

b = 237.3

If the hygristor recorded a relative humidity of 100, this would imply that the temperature in the immediate vicinity of the hygristor sensing element was the temperature resulting from the solution of the following equation:

$$100 = \left\{ e/\left[(6.11)\,(10^{at/(b+t)})\right]\right\} \times 100 \qquad (A.2)$$

This latter temperature is the dew point temperature. This will be denoted as t_D. The solution of the above equation is:

$$t_D = \frac{b\,(\log e - \log 6.11)}{a - (\log e - \log 6.11)}\,. \qquad (A.3)$$

Because vapor pressure is not affected by the temperature of the air in the immediate vicinity of the hygristor sensing element, the hygristor temperature can be related to the dew point temperature. The relationship is

$$(R/100)\left(10^{at_H/(b+t_H)}\right) = 10^{at_D/(b+t_D)} \qquad (A.4)$$

2. PROPOSED THERMISTOR TEMPERATURE CORRECTION

The temperature of the hygristor can be related to the thermistor temperature, which is regarded as an excellent estimate of the temperature of the atmosphere surrounding the rawinsonde, by the equation

$$t_H = t_T + C$$

where

t_T = Thermistor temperature

C = Correction

Thus

$$t_H/(b + t_H) = (t_T + C)/(b + t_T + C)$$

which for practical purposes can be approximated as $(t_T + C)/(b + t_T)$, since we anticipate that C will be of the order of two degrees, whereas $b + t_T$ will be from 240 to 270 degrees. Using this approximation, a solution for C is

$$\frac{C\ a}{b + t_T} = at_D/(b + t_D) - \log \left[(R/100)\ (10)^{at_T/(b + t_T)} \right] \qquad (A.5)$$

If a population of space-time points is defined such that C is essentially a constant for all points in the population, for each such space-time point, there is a triplet of observations:

- Thermistor temperature
- Hygristor relative humidity
- Dew point temperature.

Various means can be defined with reference to the population in question. The most useful means are:

$$M_1 = E\left[a/(b + t_T)\right] \tag{A. 6}$$

$$M_2 = E\left[at_D/(b + t_D)\right] \tag{A. 7}$$

$$M_3 = E\left\{\log\left[(R_H)\,(10^{at_T/(b + t_T)})\right]\right\} \tag{A. 8}$$

It follows that

$$C = (M_2 - M_3)/M_1 \tag{A. 9}$$

3. COMMENTS

(1) The correction can be estimated provided estimates of M_1, M_2, and M_3 are available. Rawinsonde data will provide estimates of M_1 and M_3 based on a random sample of observations from a defined population. Airplane data will provide an estimate of M_2 based on a second random sample from presumably the same population of space-time points.

(2) The correction process will be successful if homogeneous populations of space-time points can be defined. A population is homogeneous if the same correction essentially applies to all points.

(3) The procedure to be followed in obtaining rawinsonde humidity data for archiving or budget computations is as follows:

- Assign the rawinsonde point to an appropriate homogeneous population and hence choose the appropriate thermistor

- Estimate hygristor temperature as the sum of the thermistor temperature and the correction

- Calculate the vapor pressure by equation (A. 1)

- Calculate the dew point by equation (A. 2).

(4) It is a simple matter to modify the proposed procedure to make it unnecessary to approximate $a(t + C)/(b + t + C)$ by $a(t + C)/(b + t)$.

(5) Estimates of the standard error of the estimates of C and confidence limits for C can be easily developed by standard methods.

(6) The only property of the rawinsonde and airplane distributions that appears to have any relevance for the estimation of the temperature corrections are means. Variances, however, are useful for the construction of confidence limits.

(7) In the formation of homogeneous populations, the first step appears to be that of classifying the observations according to pressure interval, and second to classify according to location. Efforts should be made to develop a third classification according to a relative humidity-thermistor temperature index. The index is a discriminant function index that classifies observations into two subpopulations.

(8) Since separate corrections are anticipated for each of three heights, it should be possible to interpolate and extrapolate to other heights. A relationship between the correction and atmospheric density is anticipated.

(9) Detailed procedures need to be developed:

- For performing the necessary tests on the data that will tend to support or reject the above procedures
- For forming the homogeneous populations
- For making the corrections as a matter of routine
- For computing standard errors and confidence limits on the corrections.

APPENDIX B

A STATISTICAL DATA PLAN FOR BOMEX

Theodore W. Horner

Interim Report
Covering the Period
October 1969 to March 1970
prepared for
Barbados Oceanographic and Meteorological
Analysis Project (BOMAP)
Contract No. E-192-69(N), Task 4 April 10, 1970

TABLE OF CONTENTS

INDEX OF TABLES

SUMMARY

As part of the United States contribution to the Global Atmospheric Research Project (GARP), a large-scale meteorological experiment, known as the Barbados Oceanographic and Meteorological Experiment (BOMEX), was conducted near the island of Barbados in the Caribbean from May through July of 1969. The primary goal of BOMEX was to provide data for studying the sea-air interaction that drives atmospheric circulation and global weather systems. To accomplish this goal, thousands of hours of data tape and other records were collected. Twelve ships and 29 planes were utilized. More than 1,500 personnel from government agencies, private industry, and universities participated in the project. The processing and coordination of BOMEX data were the responsibility of the Barbados Oceanographic Meteorological Analysis Project (BOMAP), a project established within the Research Laboratories of the Environmental Science Service Administration.

The reduction of the BOMEX data to a form useful for archiving and for scientific computations and interpretation presents many statistical problems. The objective of the subject contract was to formulate a relevant statistical methodology. This methodology has been formulated in the following areas:

- Estimation of meteorological parameters
- Solution of nonlinear equations
- Formulation of calibration equations.

In each of the above areas a number of informal memorandums were prepared. These memorandums were discussed with the BOMAP staff and served as vehicles for emerging concepts as to appropriate statistical treatment of the data.

This report is an interim report summarizing the work under the subject contract for the first six months (October 1969 through March 1970). The work completed during the first six months has indicated the importance of very careful visual inspection and editing

of the data. The editing will require many adjustments and corrections. However, the data have a structure that will permit the application of statistical methodology and the combining of rawinsonde aircraft and dropsonde data in a meaningful way. It is now clear that the data contain usable information. However, it will not be possible to simply process most of the data. Rather, processing must proceed by a sequence of iterative cycles involving computations and summary display, interpretation, and the planning of new computations. Consequently, it is very important to leave options in processing the data; that is, any one method of analysis must not be formulated prior to examination of data and computation results.

I. INTRODUCTION

The Barbados Oceanographic and Meteorological Experiment (Project BOMEX) was the first of a series of large-scale research projects planned by many nations throughout the world under the Global Atmospheric Research Project (GARP). The primary goal of the BOMEX project was to provide data for studying the sea-air interaction that drives atmospheric circulation and global weather systems. The project was conducted from May through July of 1969 and produced substantial quantities of data that were recorded manually, on magnetic tapes, and on charts. The reduction of the raw data to a form useful for scientific investigation and for storage in scientific archives is the responsibility of the Barbados Oceanographic and Meteorological Analysis Project (BOMAP), a project established within the Research Laboratories of the Environmental Science Services Administration.

1. DATA REDUCTION AND ANALYSIS

In the reduction of the data it has been important to consider not only the meteorological theory underlying the BOMEX experiment but also the statistical properties of the data and the mechanisms and formats used in assembling the raw data. The data-reduction process presents problems with respect to filtering, averaging, and looking at the data on the proper scale. Alternative techniques, methods, and instruments were employed to estimate the same quantity, necessitating a comparison of the strengths and limitations of the various alternatives. The quantity of data varies from one source to another and from one time period to another. The data contain both systematic errors for which adjustments must be made and random errors whose magnitudes must be estimated. A high degree of autocorrelation exists in the data and this must be taken into account. The magnitude of error deviations and the conditions under which they are important require understanding and definition. Processing of the data requires the preparation of computation formulas for estimating the meteorological parameters in the most efficient manner. A knowledge of the accuracy with which the meteorological parameters are estimated will help in the design of future experiments.

2. THE ROLE OF BOOZ, ALLEN APPLIED RESEARCH INC.

Participation by Booz, Allen Applied Research Inc. (BAARINC) under the subject contract was initiated in October of 1969 following completion of the data-collection program. BAARINC was to develop a statistical methodology that would relate the BOMEX data to the end results required for budget computations in an optimum manner. This methodology requires the development of the rationale for estimation procedures and an understanding of the nature of the error in the resulting estimates.

This interim report covers the six-month period from October through March 1970. During the contract, samples of some data were available in preliminary form, but the extent of data available for inspection was strictly limited. The major portion of the data was unavailable because the analog tapes had to be digitized, transformed to engineering units, edited, and reproduced in a form suitable for study.

Because of the unavailability of the data and the frontier nature of the BOMEX experiment, it was necessary to begin the development of statistical methodology on a trial and error basis. The methodology evolved as additional data became available and as the mathematics of the meteorological budget equations were solidified.

Virtually none of this methodology had been developed in a recorded form before or during the BOMEX experiment, although Dr. Ben Davidson, the former scientific director, was probably well aware of such methodology. The untimely death of Dr. Davidson in December preceding the BOMEX experiment created a substantial gap in analysis, reduction, and interpretation procedures. This gap has had to be filled by others.

3. STUDY APPROACH AND PROJECT STATUS

The approach to the subject study during the period covered by this interim report has been dictated by two primary considerations:

- Theoretical aspects of the budget equations have been in a state of flux and are only now being formulated in a way that is useful in defining the end results of the statistical analysis.

- Only a limited amount of data were available for inspection.

The second point has prevented a detailed understanding of assumptions that can be made with reference to the data. These assumptions are necessary for formulating valid statistical analyses. For this reason we have avoided crystallizing the statistical methodology during the interim period. Rather, we have concentrated upon understanding the structure and the limitations of the data and the way in which these data can be employed in budget computations.

Informal memorandums (Table B-1) have been prepared reflecting various aspects or links in the overall analysis program. Some of the highlights of the more important memorandums are summarized in Chapters II and III of this report. Conclusions based on the work during the interim period are included in Chapter III.

The informal memorandums have been discussed with members of the BOMAP staff as a basis for an evolving methodology. This methodology has now evolved to the point where the main features of the analysis program are understood and accepted by members of the BOMAP staff. These features, which can now be brought together in a more definitive plan, will be described in future informal memorandums and in the final report on this contract.

Table B-1
List of Research Memorandums

Date	Title
10/23/69	Estimation of Error Variance for Rawinsondes (I)
11/03/69	Estimation of Error Variance for Rawinsondes (II)
12/01/69	Confidence Limits for Divergence Using Aircraft Data
12/05/69	Solution of Nonlinear Equations
01/09/70	Estimation of Volume Parameters of the BOMEX Box
01/29/70	Analysis of Bias Corrections for Calibration Boxes
02/05/70	Employment of Dropsonde Observations
02/16/70	Dropsonde Observations and the Linear x Linear Term
02/20/70	Estimation of Scalar Parameters
02/23/70	Simplification of the Calibration Equations
03/06/70	Estimation of Scalar Parameters II

II. ESTIMATION OF BUDGET PARAMETERS

A central objective of the BOMEX experiment was to provide data useful for studying the flux of energy from the ocean to the atmosphere. This task is to be accomplished by keeping budgets on the passage into and out of the BOMEX box* on the following properties:

- Mass
- Momentum-zonal
- Momentum-meridional
- Enthalpy
- Mechanical energy
- Total energy.

The meteorological parameters employed in budget calculations divide, from the standpoint of estimation, into two types:

- Scalar parameters
- Divergence parameters.

Primary emphasis during the subject contract was placed on the development of procedures for the estimation of scalar parameters. Ideally, such estimation makes use of information at points inside the BOMEX box, although estimation is still possible using only points on the sides of the box. Moreover, the procedures are applicable, though to a lesser extent, to the estimation of some divergence parameters. Divergence parameters that make use of the divergence theorem are most efficiently estimated, since they require information only at points on the surface of the box.

---

* The BOMEX box is located in an ocean area near the island of Barbados in the Caribbean. It is roughly 500 kilometers square and 500 millibars of pressure differential high. Extensive meteorological observations were collected around and within the box by means of ships, airplanes, satellites, and buoys.

Scalar quantities, such as temperature and relative humidity, have a conceptual numerical value at every point in and around the BOMEX box. The scalar may also change with time.

Interest centers in an average of the scalar over a working box located at a height of p^* where p^* is the millibar difference between atmospheric pressure at surface level and at a height z above the surface. A working box is a horizontal layer of the Bomex box covering a selected pressure differential. The working box will be assumed to be sufficiently thin so that it can be treated as if it were a plane. It should be wide enough, however, to obtain a sufficient quantity of observations so that meteorological quantities associated with the working box can be estimated with reasonable precision. Subsequent calculations required for budget computations and archiving will require combining the summary statistics over several adjacent working boxes to cover a 25-millibar pressure differential or the entire BOMEX box.

In order to indicate the general nature of the estimation procedure, we will first define a coordinate system. This definition will permit formal mathematical definition of the properties of the working box that are to be estimated. The primary properties are the mean and the variance of the scalar. A number of scalars, such as temperature, relative humidity, divergence, and kinetic energy, are of interest.

Observations are available from the BOMEX experiment for estimating the properties of the working box. If the observations had been distributed at random throughout the working box, then a simple sample mean and variance would constitute unbiased estimates of the true mean and variance of the working box. The observations, for excellent reasons, are not random; thus, a method of estimation is required that takes into account the structure of the observations. The observations were taken primarily at the corners of the BOMEX box, along the sides, and along diagonals in the interior of the working box. Observations from all three sources are present for only part of the working boxes.

A data structure has meaning only with reference to a function that characterizes the variation of the scalar over the working box; that is, a set of observations may have a very good structure with respect to one function, but that same structure may be poor with respect to a second function. The functional forms of scalars are unknown, although they can be approximated in various ways. The proposed method of estimation is to relate the structure of the data to a mutually-agreed-upon approximation function.

A Taylor's expansion is one device for obtaining a useful approximation function. The approximation lends itself to the application of standard statistical methodology. Thus, using a Taylor's approximation, each observation can be expressed as a linear function of the parameters of the Taylor's expansion. The least square procedures required to estimate the parameters in such linear function are well known. A number of computer packages are available for accomplishing such estimations. For this reason, the computational procedures for estimating the parameters of the linear Taylor's expansion have not been emphasized in this report.

In the sections that follow, we have illustrated the general procedure of employing approximation models by using a Taylor's approximation through the quadratic terms. Although the quality of the data may not permit the estimation of the quadratic terms of the approximation, the subsequent derivations should have value in indicating the informational loss that accrues from using a simpler approximation model.

1. COORDINATE SYSTEM

A rectangular coordinate system will be defined with axes parallel to the sides of the BOMEX box. It is convenient to refer to the two axes as east-west (plus direction toward the east) and north-south (plus direction toward the north), although these axes do not point exactly in the indicated directions because of a slight counterclockwise rotation of the BOMEX box. The east-west axis will be referred to as the X_1 axis, and the north-south as the X_2 axis. The center of the BOMEX box is located at (X_{1c}, X_{2c}). The values of X_1 and X_2 at the center and for particular observations are obtained from functions of the longitude and latitude of the observations.

It will be convenient to define new coordinates x_1 and x_2 as shown below:

$$x_1 = \frac{X_1 - X_{1c}}{L/2}$$

$$x_2 = \frac{X_2 - X_{2c}}{L/2} \qquad \text{(B. 1)}$$

where L is the length of a side of the BOMEX box measured in the same units as X_1 and X_2.

2. DEFINITION OF QUANTITIES TO BE ESTIMATED

A meteorological quantity, such as the average temperature of a working box, should be distinguished from the corresponding observed-point quantity, such as the temperature at a particular point in the working box. A derived-point quantity is the function of one or more observed point quantities. A derived-point quantity will be represented by F(u, v) where u and v are scalar functions of x_1 and x_2. A volume quantity is defined as a function of a derived-point quantity over all points of the volume under consideration.

The volume quantity of most interest is the mean taken over a working box. This mean can be represented as

$$F(u, v) = (1/4) \int_{-1}^{1} \int_{-1}^{1} F(u, v)\, dx_1\, dx_2 \qquad \text{(B. 2)}$$

In some instances the variance of the derived-point quantity over the working box is important. This variance is

$$V_F = (1/4) \int_{-1}^{1} \int_{-1}^{1} [F(u, v) - \overline{F}(u, v)]^2\, dx_1\, dx_2 \qquad \text{(B. 3)}$$

Forms for the function F(u, v) that are of particular interest are listed in Table B-2.

Table B-2
Typical Forms for F(u, v)

Form	Comments
$F = u$	where u is a scalar
$F = u^2$	
$F = uv$	where u and v are two scalars
$F = \frac{\delta u}{\delta x_1} + \frac{\delta v}{\delta x_2}$	where u is the velocity in the x_1 direction and v is the velocity in the x_2 direction at x_1, x_2; that is, the divergence at x_1, x_2
$F = u^2 + v^2$	where u is the velocity in the x_1 direction and v_2 is the velocity in the x_2 direction. F is twice the kinetic energy at x_1, x_2.

3. APPROXIMATION FUNCTIONS

Scalar properties u and v are functions of x_1 and x_2. Generally, the forms of these functions are not known. However, the functions can be approximated as a Taylor's expansion around $x_1 = x_2 = 0$. The accuracy and validity of the approximation improves as more terms are included in the Taylor's expansion. The foregoing statement is valid with respect to the true but unknown Taylor's expansion. The statement does not necessarily hold for the approximation to the Taylor's expansion that must be employed in practice.

The latter is derived from estimates, which are derived from data, for the true coefficients of the Taylor's expansion. The properties of the estimates depend upon the structure and quantity of the available data. An approximation to the Taylor's expansion with n_1 terms may not be as good as an approximation to the true function u as another approximation with n_2 terms even though n_2 is smaller than n_1.

An approximation u_a for u using Taylor's expansion through the quadratic terms can be written as

$$u_a = B_0 + B_1 x_1 + B_2 x_2 + B_{11} x_1^2 + B_{12} x_1 x_2 + B_{22} x_2^2 \qquad (B.4)$$

where B_0 is the value of u for $x_1 = x_2 = 0$. A Taylor's expansion for v can be written in a similar manner. Primes will be placed on the B's to differentiate them from those for u.

4. APPROXIMATION FUNCTIONS AND MEAN OF THE WORKING BOX

The functions u and v can be written as

$$u = u_a + e_u \quad \text{and} \quad v = v_a + e_v \qquad (B.5)$$

where e_u and e_v are error deviations that are a function of x_1 and x_2. The approximation function F(u, v) can be written as

$$F(u, v) = F(u_a + e_u,\ v + e_v)$$

or (B.6)

$$F(u, v) = F(u_a, v_a) + e_F$$

where e_F is an error deviation, again a function of x_1 and x_2.

The mean of F(u, v) over the working box can be written as

$$\bar{F}(u, v) = \bar{F}(u_a, v_a) + \bar{e}_F \tag{B.7}$$

Since the mean of the working box F(u, v) will be estimated by the mean of the approximation function $F(u_a, v_a)$, the accuracy of the estimate will be affected by two sources of error:

- Approximation error (e_F)

- Sampling error associated with the estimate of the mean of the approximation function over the working box.

As more terms are employed in the Taylor approximation function, the approximation error tends to be reduced. The sampling error of the quadratic and higher-order terms of the approximation function tends, however, to be much larger than the linear terms of the expansion. Thus, a greater quantity and quality of data are required to estimate the nonlinear terms with useful precision.

5. APPROXIMATION FUNCTION AND VARIANCE OF THE WORKING BOX

For brevity F(u, v) and $F(u_a, v_a)$ will be represented as F and F_a. Since $F - \bar{F}$ can be written as

$$F - \bar{F} = F_a + e_F - \bar{F}_a - \bar{e}_F = (F_a - \bar{F}_a) - (e_F - \bar{e}_F) \tag{B.8}$$

the variance of F over the working box can be written as

$$V_F = V_{F_a} + \mathrm{Cov}(F_a, e_F) + V_{e_F} \tag{B.9}$$

where V_{F_a} and V_{e_F} are the variances of F_a and e_a and $\mathrm{Cov}(F_a, e_F)$ is the covariance of F_a and e_F.

To the extent that u_a and v_a are good approximations of u and v, the terms Cov (F_a, e_F) and V_{e_F} will be small relative to V_{F_a}, and V_{F_a} should constitute a good approximation to V_F. Unfortunately, it is anticipated that the BOMEX data will contain little information as to the magnitudes of e_u and e_v, and hence little information as to the magnitude of the approximation error.

6. ESTIMATION OF VOLUME QUANTITIES

The definition of the mean and variance of the scalar over the working box were in terms of integrals of the approximation function, a function linear in its parameters. The integration of the approximation function can be accomplished term by term, so the mean of the working box can also be expressed as a function linear in the parameters of the approximation function. The coefficients of the parameters are the result of the term-by-term integrations.

All of the integrations are similar, since they are integrations of $x_1^r x_2^s$ over the x_1, x_2 plane. The various terms are distinguished by the values assigned to r and s. The integration of the general term $x_1^r x_2^s$ is given below:

$$G(r, s) = (1/4) \int_{-1}^{1} \int_{-1}^{1} x_1^r \, x_2^s \, dx_1 \, dx_2 = [1/(r+1)]\,[1/(s+1)] \; \delta_r \; \delta_s \quad (B.10)$$

where

$$\delta_r = \begin{cases} 1 \text{ for r even} \\ 0 \text{ for r odd} \end{cases} \qquad \delta_s = \begin{cases} 1 \text{ for s even} \\ 0 \text{ for s odd} \end{cases}$$

It is evident from (B. 10) that terms with either r or s odd vanish. Once F_a or $(F_a - \bar{F}_a)^2$ is written out, the use of (B. 10) enables one to immediately write out the volume quantity. Specific examples of the use of (B. 10) for the functions given in Table B-2 are developed in the subsections that follow. Corresponding results for approximation functions employing only linear terms are immediately derived by dropping off the quadratic terms.

7. FORM F = u

The first form in Table B-2 was F = u. The approximation function (B.4) will be used in place of u. Integration is performed term by term. The values of r and s for the several terms of F are set out below.

Term of F	r	s	r	s	G(r, s)
B_o	0	0	1	1	1
$B_1 x_1$	1	0	0	1	0
$B_2 x_2$	0	1	1	0	0
$B_{11} x_1^2$	2	0	1	1	1/3
$B_{12} x_1 x_2$	1	1	0	0	0
$B_{22} x_2^2$	0	2	0	1	1/3

The value of $\overline{F}$ is

$$F = B_o\, G(0,0) + B_1 G(1,0) + B_2 G(0,1)$$

$$+ B_{11}\, G(2,0) + B_{12} G(1,1) + B_{22} G(0,2)$$

$$= B_o + (1/3)(B_{11} + B_{22}).$$

The estimate of $\overline{F}$ is

$$\hat{\overline{F}} = \hat{B}_o + (1/3)(\hat{B}_{11} + \hat{B}_{22}). \qquad \text{(B. 11)}$$

The estimated variance of the estimate is

$$\sigma^2_{\hat{\overline{F}}} = \hat{V}_o + (1/9)(\hat{V}_{11} + \hat{V}_{22})$$

$$+ (2/3)\,[\hat{\text{cov}}\,(\hat{V}_o, \hat{V}_{11}) + \hat{\text{cov}}\,(V_o, \hat{V}_{22})]$$

$$+ (2/9)\,[\hat{\text{cov}}\,(\hat{V}_{11}, \hat{V}_{22})]. \qquad \text{(B. 12)}$$

In the above formula, ^ (read hat) placed over a symbol indicates an estimate of the quantity represented by the symbol. The subscripts of the V's reference the quantity to which the variance relates. Thus, $\hat{V}_{11}$ is the estimated variance of the estimate of B_{11}. The quantity cov $(\hat{V}_o, \hat{V}_{22})$ is the estimated covariance of the estimates of $\hat{V}_o$ and $\hat{V}_{22}$. The estimates of the coefficients, the variances, and the covariances are obtained by least squares procedures.

8. FORM F=uv

The solution for $\bar{F}$ is facilitated through the use of Table B-3. Table entries are associated with the 6 x 6 = 36 products that can be formed by taking each term of u with each term of v. A typical term in u is the triple product involving a B coefficient, a power i of x_1 and a power of x_2; that is, the typical term is

$$Bx_1^i x_2^j$$

Similarly, the typical term in v is $B' x_1^{i'} x_2^{j'}$. The typical product term is

$$B\,B'\,x_1^i x_2^j x_1^{i'} x_2^{j'}.$$

The values of r and s in the product term are

$$r = i + i'$$

and

$$s = j + j'.$$

Table B-3 catalogs combinations of r and s that are associated with the 36 product terms. For example, consider the term $B_{11}\,x_1^2$ in u and $B_{12}\,x_1\,x_2$ in v. In this case

$$i = 2,\ j = 0,\ i' = 1,\ j' = 1.$$

Table B-3
Values of r and s in the Function F = uv

			Parameters in v					
			B'_o	B'_1	B'_2	B'_{11}	B'_{12}	B'_{22}
		Values of i' in u	0	1	0	2	1	0
		Values of j' in v	0	0	1	0	1	2
Parameters in u	Values of i in u	Values of j in v						
B_o	0	0	(0, 0)*	(1, 0)	(0, 1)	(2, 0)*	(1, 1)	(0, 2)*
B_1	1	0	(1, 0)	(2, 0)*	(1, 1)	(3, 0)	(2, 1)	(1, 2)
B_2	0	1	(0, 1)	(1, 1)	(0, 2)*	(2, 1)	(1, 2)	(0, 3)
B_{11}	2	0	(2, 0)*	(3, 0)	(2, 1)	(4, 0)*	(3, 1)	(2, 2)*
B_{12}	1	1	(1, 1)	(2, 1)	(1, 2)	(3, 1)	(2, 2)*	(1, 3)
B_{22}	0	2	(0, 2)*	(1, 2)	(0, 3)	(2, 2)*	(1, 3)	(0, 4)*

NOTES: The table entry is (r, s).

Terms followed by an asterisk (*) do not vanish; that is, $\delta_r = \delta_r = 1$.

It follows that

$$r = 2 + 1 = 3, \delta_r = 0$$

and

$$s = 0 + 1 = 1, \delta_s = 1.$$

Since G(3, 1) = (1/4)(1/2)(0)(1) = 0, the term in question does not enter into the product of u and v.

Using 3, the average of F over the working box is readily seen to be

$$\begin{aligned} F = B_o B_o' &+ (1/3)\,[B_o(B_{11}' + B_{22}') + B_o'(B_{11} + B_{12}) + B_1 B_1' + B_2 B_2'] \\ &+ (1/5)(B_{11}B_{11}' + B_{22}B_{22}') \\ &+ (1/9)(B_{22}B_{11}' + B_{12}B_{12}' + B_{11}B_{22}'). \qquad \text{(B. 13)} \end{aligned}$$

9. FORM $F=u^2$

The form $F=u^2$ can be treated as a special case of F=uv by letting v=u. The formula for $\overline{F}$ is

$$\begin{aligned} \overline{F} = B_o^2 &+ (1/3)\,[2B_o(B_{11} + B_{22}) + B_1^2 + B_2^2] \\ &+ (1/5)\,(B_{11}^2 + B_{22}^2) \\ &+ (1/9)\,(2B_{11}B_{22} + B_{12}^2). \qquad \text{(B. 14)} \end{aligned}$$

10. FORM $F=u^2 + v^2$

The formula for F is

$$\bar{F} = B_o^2 + (1/3)\,[2B_o(B_{11} + B_{22}) + B_1^2 + B_2^2]$$
$$+ (1/5)\,[B_{11}^2 + B_{22}^2]$$
$$+ (1/9)(2B_{11}B_{22} + B_{12}^2)$$
$$+ (B'_o)^2 + (1/3)\,[2B'_o(B'_{11} + B'_{22}) + (B'_1)^2 + (B'_2)^2]$$
$$+ (1/5)\,[(B'_{11})^2 + (B'_{22})^2]$$
$$+ (1/9)\,[2B'_{11}B'_{22} + (B'_{12})^2]. \qquad \text{(B. 15)}$$

11. FORM $F = \frac{\partial u}{\partial x_1} + \frac{\partial v}{\partial x_2}$

The function F reduces to

$$\bar{F} = B_1 + 2B_{11}x_1 + B_{12}x_2 + B'_2 + B'_{12}x_1 + 2B'_{22}x_2 \qquad \text{(B. 16)}$$

The mean over the working box is

$$\bar{F} = B_1 + B'_2$$

12. ESTIMATES AND VARIANCES OF ESTIMATES

F is estimated by substituting estimates of the B's that are derived from data. The observations used in estimating the B's may not all occur at the same time. Since the scalar may vary as a function of time, a decision must be made as to the manner of handling the effects of time differences in the observations. If the time differences are small, then it should be satisfactory to simply treat all observations as if they occurred at the same time. Another method is to include time terms in the linear model for the observations. This would effectively adjust the estimates of the B's to discount the effects of time variations. Estimates of the B's should be accompanied by estimates of their variances and covariances.

F has the form $F = \sum_i D_i T_i$ (B.17)

where D_i is the coefficient of a term T_i that is some function of the B's. The variance of F is

$$V_F = \sum_i D_i^2 V_i + 2 \sum_{\substack{i \\ i' >}} \sum_{\substack{i \\ i}} D_i D_{i'} \operatorname{Cov}(T_i, T_{i'}) \tag{B.18}$$

where V_i = the variance of T_i

and Cov $(T_i, T_{i'})$ = the covariance of T_i and $T_{i'}$.

13. VARIANCE OF F=u OVER A WORKING BOX

The variance of the function F over the working box was defined as

$$V_F = (1/4) \int_{-1}^{1} \int_{-1}^{1} [F(u,v) - \bar{F}(u,v)]^2 \, dx_1 dx_2$$

This equation can also be written in the form

$$V_F = (1/4) \int_{-1}^{1} \int_{-1}^{1} [F(u,v)]^2 \, dx_1 dx_2 - [\bar{F}(u,v)]^2.$$

For F=u, the square of F is

$$\bar{F}^2 = B_o^2 + (2/3) B_o (B_{11} + B_{22}) + (1/9)(B_{11}^2 + 2B_{11}B_{22} + B_{22}^2)$$

Using equation (B.14), the variance of the function F over the working box is derived as

$$V_F = (1/3)(B_1^2 + B_2^2) + (4/45)(B_{11}^2 + B_{22}^2) + (1/9) B_{12}^2 \tag{B.19}$$

14. VARIANCE OF $F = \frac{\partial u}{\partial x_1} + \frac{\partial v}{\partial x_2}$ OVER A WORKING BOX

The function $F-\overline{F}$ can be written as

$$F - \overline{F} = 2B_{11}x_1 + B_{12}x_2 + B'_{12}x_1 + 2B'_{22}x_2$$

$$= (2B_{11} + B'_{12})\, x_1 + (B_{12} + 2B'_{22})\, x_2$$

The variance over the working box is

$$V_F = (2B_{11} + B'_{12})^2\, G(2,0)$$

$$+ 2(2B_{11} + B'_{12})(B_{12} + 2B'_{22})\, G(1,1)$$

$$+ (B_{12} + 2B'_{22})^2\, G(0,2)$$

$$= (1/3)\,[(2B_{11} + B'_{12})^2 + (B_{12} + 2B_{22})^2].$$

III. OTHER RESEARCH TOPICS

Although the principal research effort of the contract was associated with the statistical methodology for the budget computations, research was also directed to two additional areas:

- Confidence limits for divergence using aircraft data
- Solution of nonlinear equations.

1. CONFIDENCE LIMITS FOR DIVERGENCE USING AIRCRAFT DATA

During four closely-spaced nights of the BOMEX experiment, an aircraft flew the perimeter of the BOMEX box collecting data on the normal component of the wind. These data were employed to compute rough preliminary estimates of divergence of a working box at the level at which the aircraft flew. The estimation procedure was corrected for bias in the normal wind component by making use of calibration data taken at seven heights on each side. At each height, a difference in velocity was computed by subtracting the normal velocity in the clockwise direction from the velocity in the counterclockwise direction. One-half of the average of these differences over the several heights for a side was used as a correction to the average normal velocity for the side.

A methodology for computing confidence limits for estimates of the divergence was developed. The estimate of divergence for a given night was calculated from the formula

$$\hat{D} = (1/L) \sum_{i=1}^{4} (\bar{v}_i - (1/2)\,\bar{d}_i)$$

where $\hat{D}$ = the divergence estimate

L = the length of a side of the BOMEX box

$\bar{v}_i$ = the average of the observed normal velocities for side i

$\bar{d}_i$ = the average difference between the normal velocities in the clockwise and counterclockwise directions for the calibration legs of side i.

D was treated as a normal deviate in computing 95-percent confidence limits. The standard error of D was derived as

$$s_{\hat{D}}^2 = (1/L^2) \left\{ \sum [s_{\bar{v}_i}^2 + (1/4)s_{\bar{d}_i}^2] \right\}$$

where $s_{\bar{v}_i}$ and $s_{\bar{d}_i}$ are the estimated standard errors of $\bar{v}_i$ and $\bar{d}_i$, respectively.

2. SOLUTION OF NONLINEAR EQUATIONS

Wind-velocity data collected as an aircraft flew around the perimeter of the BOMEX (box) require certain bias corrections, and each side of the box required a different set of corrections. The corrections are calculated from data obtained by flying a small square at each corner of the BOMEX box. The square was always flown in a counterclockwise direction.

The four sides of the square give rise to eight equations for computing corrections. The two equations on a given side are associated respectively with east-west and north-south wind velocity. Since the eight equations have eight unknowns, a unique solution for the unknowns exists. This solution, however, is complicated by the nonlinearity of the equations in the parameters.

Since a separate set of corrections is required for each side of the BOMEX box on each flight, and since there were many such flights, a large number of solutions will be required. For this reason, it is worthwhile to develop a rapid and efficient method of solution that would lend itself to computerization.

On the i-th side (i = 1, 2, 3, 4) of each square, the following quantities were measured:

GSM_i = measured groundspeed

$TASM_i$ = measured true airspeed

$hdgM_i$ = measured aircraft heading

δM_i = measured drift angle

The eight unknown variables were:

GE = groundspeed error

TE = true airspeed error

δE_i = drift-angle error (heading-dependent), i=1, 2, 3, 4

VXA = wind velocity west

VYA = wind velocity south

The equations relating the observables to the unknowns, as prepared by Robert W. Reeves of the BOMAP staff, are set out below.

$$\begin{aligned} VXA = {} & GSM_i \sin(hdgM_i + \delta M_i) - TASM_i \sin(hdgM_i) \\ & + GE \sin(hdgM_i + \delta M_i) - TE \sin(hdgM_i) \\ & + \delta E_i\, GSM_i \cos(hdgM_i + \delta M_i) + \delta E_i\, GE \cos(hdgM_i + \delta M_i) \end{aligned} \qquad (B.20)$$

$$\begin{aligned} VYA = {} & GSM_i \cos(hdgM_i + \delta M_i) - TASM_i \cos(hdgM_i) \\ & + GE \cos(hdgM_i + \delta M_i) - TE \cos(hdgM_i) \\ & - \delta E_i\, GSM_i \sin(hdgM_i + \delta M_i) - \delta E_i\, GE \sin(hdgM_i + \delta M_i) \end{aligned} \qquad (B.21)$$

A solution to these equations was developed by a process of transforming variables and searching for the value of a parameter that reduces a least squares error sum of squares to zero. The methodology lends itself to computerization and a computer program has been written that reflects the methodology.

IV. CONCLUSIONS

During the interim period covered by this report, a number of conclusions were developed with respect to statistical treatment of the BOMEX data. These conclusions are given below.

(1) The quantities required for budget calculations are quantities associated with working boxes. These quantities are horizontal slices of the BOMEX box with a thickness of a 25-millibar pressure differential. Estimation of budget quantities can be accomplished by combining estimates associated with working boxes.

(2) Two types of estimates are required for the working boxes:

- The average of a quantity defined at every point in the working box, for example, the average temperature.

- The average over all points of the four vertical sides of the working box; for example, the product of wind velocity normal to the face and relative humidity.

(3) The nature of the errors, to which the estimates of the working box statistics are subject, is amenable to statistical analysis and can be expressed in the form of confidence limits for the estimated parameters.

(4) Before BOMEX data can be employed for scientific computations, substantial editing, corrections, and adjustments must be made.

(5) The nature of the BOMEX data dictates a cyclic approach to data analysis, since computations must be performed, summarized, displayed, interpreted, and used as a basis for planning subsequent computations.

(6) The most effective method of rapidly cycling computations that reflect statistical analysis of the data is the formation of scientific building blocks. These blocks will contain the summary statistics in the form of numbers of observations, means, variances, and covariances. The blocks will permit rapid combination of the foregoing statistics in alternative patterns, thereby permitting the selection of an optimum pattern based on the characteristics of the data.

(7) The BOMEX data contain information useful for budget computations that can be extracted by appropriate statistical analysis; that is, the "noise" in the data is not so excessive that the "signal" will be completely obscured. The signal-to-noise ratio is sufficiently low, however, to require substantial effort to elicit the informational content.

☆ U. S. GOVERNMENT PRINTING OFFICE : 1971 431-994/177