EPA-650/4-75-005

OBJECTIVE PROCEDURES FOR OPTIMUM LOCATION OF AIR POLLUTION OBSERVATION STATIONS

by

C. Eugene Buell

Kaman Sciences Corporation
1500 Garden of the Gods Road
P. O. Box 7463
Colorado Springs, Colorado 80933

Contract No. 68-02-0699

EPA Project Officer: Kenneth L. Calder

Meteorology and Assessment Division
Environmental Sciences Research Laboratory
Research Triangle Park, North Carolina 27711

Prepared for

U. S. ENVIRONMENTAL PROTECTION AGENCY
Office of Research and Development
Environmental Sciences Research Laboratory
Research Triangle Park, North Carolina 27711

June 1975

DISCLAIMER

This report has been reviewed by the Environmental Sciences Research Laboratory, U.S. Environmental Protection Agency, and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the U.S. Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

ACKNOWLEDGMENTS

We express our appreciation for the assistance of Mrs. Barbara Byerly for basic programming and especially to Mr. Edwin Bathke for invaluable assistance in solving our major problems and preparing this report.

ABSTRACT

This document is concerned with the development of linear regression techniques for interpolation of air pollutant concentrations over an area and, using these techniques, the construction of a computer program for determining the optimum location of air pollution observing stations. The general interpolation problem is surveyed in the first chapter and the advantages of using linear regression formulas as interpolation formulas are discussed. Special emphasis is placed on the case in which the observations contain errors of observation or effects of limited range of influence. Since the use of linear regression methods depends on knowledge of the two-point correlation function for pollutant concentration measures, the construction of correlation coefficient functions from synthetic data is taken up, together with methods for interpolation of this information. Considerable attention is given to the estimation of residual variances or the effects of limited range of influence. It is pointed out that certain aspects of Factor Analysis can be used for this purpose. These methods are extended to a continuous formulation of the problem in integral equation form and it is shown that the lack of accuracy in the strictly mathematical process of solution of the integral equation tends to be more important than the statistical significance of the data unless the residual variances are removed. If this is done, then the tests for accuracy and statistical significance are reconciled.

The computer program depends heavily on this last point. It appears to work well when the residual variances are carefully handled. Many of the difficulties encountered in this program were traced to this source so that users of this program should be aware of this in constructing the input materials. The reader's attention is directed to Chapter IV where this is discussed in detail.

PREFACE

The spatial distribution of air quality over an urban area represents a statistically random field of pollutant concentration since, even in principle, it is not possible to specify its structure in a deterministic fashion, but only in terms of various statistical properties. The space- and time-variability of air quality thus present difficult questions that are very inadequately and simplistically analyzed at the present time, and of which a fundamental understanding is still lacking. The accuracy of any analysis that utilizes air quality data evidentally depends on the intrinsic accuracy of the data and the density of the sampling network at which the data are available, since based on the latter values, it will normally be necessary to interpolate values for intermediate locations. An analysis of the criteria for objective selection, i.e., that does not involve personal or subjective judgment, of the optimum sampling network does not exist at the present time and is urgently needed. An "optimum network" is here meant in the sense of a network that is free from redundant observations, namely, data that could be derived with "sufficient accuracy" by some specified interpolation procedures from the given sampling network. Similar decisions are required in establishing meteorological observation networks and in this case have received a great deal of attention that is reported in an extensive modern literature. It is now required to develop and extend the appropriate statistical methodology so that it will be directly applicable to the selection of air quality sampling networks, and having due regard to both the cost and informational content of the network. At the time the present study was initiated this appeared to be of particular importance in view of the then forthcoming EPA Regional Air Pollution Study which required the establishment of both a large air quality sampling network and also an extensive meteorological network.

Very early in the research described in this report the exceptional subtleties and sophisticated difficulties of the optimization problem became apparent, and an unexpectedly large number of unforseen statistical, mathematical and computational problems were uncovered. Several different approaches, even involving variations of the logical structure of the problem, were explored by the contractor in a highly innovative fashion. Because of time limitations it was finally decided to specialize the problem and to study in some detail a one-step-at-a-time add-on method of locating sampling stations. For this it is assumed to start with that there are a few existing observation points, or at least a few points at which observations will be made based on prior considerations. On the basis of this starting network of observation points, a procedure is then developed to determine the location where the statistical error of estimate, using a simple linear-type interpolation formula in terms of the observed network concentration values, would be largest. This point would then be accepted as a best location for a new observation point, and the process repeated in an iterative fashion until the required number of station locations had been determined.

Unfortunately, a fully operational and purely objective computerized program was not achieved within the scope of the present contract. However, in support of further study of the problem, it is considered very desirable to make readily available a complete account of the present research. This is now offered in the hopes of it becoming a major contribution towards future resolution of an exceptionally difficult and subtle problem that continues to be of major importance in defining the spatial distribution of air quality.

Research Triangle Park,
North Carolina
December 1975

Kenneth L. Calder
Chief Scientist
Meteorology & Assessment Division
Environmental Sciences Research Lab.

TABLE OF CONTENTS

	Page
ABSTRACT	iii
PREFACE	iv
TABLE OF CONTENTS	vi
CHAPTER I. INTERPOLATION OF POLLUTANT CONCENTRATIONS	1
A. Interpolation Between Observation Points	1
B. Linear Regression as an Interpolation Formula	6
1. Derivation of the Basic Relations	6
2. Interpretation of Results	9
a. The point P at a point of P_i	9
b. Continuous Correlation Coefficient	10
c. Extremely Discontinuous Correlation Coefficient	11
d. Discontinuous Correlation Coefficient	12
e. Range of Influence	13
f. Small Scale Effects	14
g. Summary	15
3. Error of Interpolated Values	16
a. Implications for Measurement Locations	17
CHAPTER II. DETAILED FORMULATION OF THE PROBLEM	19
A. Area Covered	19
B. Observation Collection	20
C. Source Locations, Types, Rates	21
D. Meteorological Conditions	21
E. Type of Pollution Measurements	22
F. Interpolation/Extrapolation Methods	22
G. Optimization Method	28

TABLE OF CONTENTS (Cont'd)

Page

CHAPTER III. THE CORRELATION OF POLLUTANT CONCENTRATION BETWEEN POINTS 30

A. The Model Used to Generate Synthetic Data 30

B. Examples of the Correlation Coefficients 30

C. Interpolation of the Correlation Coefficients 36

D. Generation of Synthetic Pollution Correlation Coefficients 42

1. Main Program 42

2. Subroutines 43

CHAPTER IV. THE ANALYSIS OF THE COVARIANCE MATRICES 56

A. Evaluation of the Effect of Errors of Measurement or of Small Scale Phenomena 56

B. Determination of the Residual Variances 61

1. Representation of a Matrix in Terms of Proper Functions/Proper Values 62

2. Factor Analysis Methods 64

3. Integral Equation Methods 70

a. The Jump Discontinuity 72

b. The Quadrature Factors 74

c. The Product Integral Technique 85

d. Test for Accuracy of Solutions 88

C. Analysis of St. Louis SO_2 Data 91

1. The Basic Data 91

2. Quadrature Factors and Techniques 98

3. Factor Analysis Methods 111

D. Summary and Conclusions 123

CHAPTER V. PROGRAM BAST AND ITS SUBROUTINES 125

A. Program BAST 126

B. Subroutine CIRCUM 148

C. Subroutine FMFP 151

D. Subroutine FUNB 158

E. Subroutine FUNCT 161

TABLE OF CONTENTS (Cont'd)

		Page
F.	Subroutine CORFUN	167
G.	Subroutine INT2D	173
H.	Subroutine MATINV	175
I.	Subroutine ADDPT(XO,YO)	178
J.	Subroutine TRIFIX	184
K.	Subroutine PORDR	188
L.	Subroutine AR2	190
M.	Subroutine TRITST	192
REFERENCES		194
APPENDIX A:	Conditions on an Empirical Formula for a Correlation Coefficient	A-1
APPENDIX B:	Note on the Rank of a Covariance Matrix	B-1

CHAPTER I

INTERPOLATION OF POLLUTANT CONCENTRATIONS

The relationships between the correlation function of pollutant concentrations at two points and the spacing of observation points are discussed in the first two sections. The first section is devoted to a simple analysis of the error of interpolation using standard interpolation methods. This serves to introduce some of the basic ideas involved. The next section is devoted to the use of a linear regression as an optimum interpolation formula. Several important aspects of the problem are covered, particularly the role of errors and small scale effects. The use of the linear regression as an interpolation formula is intimately connected with the idea of a Wiener Filter (Wiener, 1949) for both smoothing and interpolation.

A. Interpolation Between Observation Points

The pollutant concentration is observed at a network of points P_i and has values x_i at these points. The total field of pollutant concentration is depicted by drawing contours of equal concentration values which are determined by the observed values. Usually, the values are sketched "by eye". To make the procedure quantitative for analysis purposes, it is necessary to specify a particular procedure used to determing the interpolated pollution field. This is done by specifying an interpolation formula that is used to describe the process. Examples of such formulas are given in Abramowitz & Stegun, 1964, p. 882. Linear interpolation between two points:

$$f(x_o+ph) = (1-p)f_{o,o} + pf_{1,o}$$

Three point formula (plane fitted to the data)

$$f(x_o+ph, y_o+qk) = (1-p-q)f_{o,o} + pf_{1,o} + qf_{o,1}$$

Four point formula (hyperbolic paraboloid fitted to the data)

$$f(x_o+ph, y_o+qk) = (1-p)(1-q)f_{o,o} + p(1-q)f_{1,o}$$
$$+ q(1-p)f_{o,1} + pqf_{1,1}$$

Six point formula (general quadratic fit)

$$f(x_o+ph, y_o+qk) = [q(q-1)/2]f_{o,-1} + [p(p-1)/2]f_{-1,o}$$
$$+ (1 + pq - p^2 - q^2)f_{o,o} + [p(p - 2q + 1)/2]f_{1,o}$$
$$+ [q(q - 2p + 1)/2]f_{o,1} + pqf_{1,1}$$

In the above formulas $f_{a,b}$ represents the concentration of the pollutant at the point with coordinates x=a,y=b, with the point (x_o,y_o) taken as (0,0). The points (x,y) are on a rectangular grid spaced h and k units apart in the x and y directions respectively. The parameters (p,q) are usually confined to the range (0,1), but not necessarily so depending on the formula.

The efficiency of the interpolation formula may be evaluated by estimating the error that would occur. We use the simple linear interpolation formula as an example to keep the arithmetic within bounds and because it illustrates the essential features of the problem. The mean square difference between the actual value and the interpolated value may be written as

$$\overline{E^2} = \overline{(y-(1-p)x_0-px_1)^2}$$

where the over-bar indicates a suitable average value.

Expanding this, and assuming that the mean field has been removed so that x_0, x_1, and y are departures from the mean, then,

$$\overline{E^2} = \overline{y^2} - 2(1-p)\overline{x_0 y} - 2p\,\overline{x_1 y} + (1-p)^2\overline{x_0{}^2} + 2p(1-p)\overline{x_0 x_1} + p^2\overline{x^2{}_1}$$

To further simplify the situation, assume that the concentration variances at P, where y is measured (the interpolated coordinate) and at P_0 where x_0 is measured and P_1 where x_1 is measured are all the same. We denote this common variance by the symbol σ^2. Then note that

$$\overline{x_0 y} = \sigma^2 r(y,x_0)$$

$$\overline{x_1 y} = \sigma^2 r(y,x_1)$$

$$\overline{x_0 x_1} = \sigma^2 r(x_0,x_1)$$

where r(a,b) is the ordinary correlation coefficient relating concentrations at the points A and B, where a and b are measured, respectively. Then

$$\overline{E^2} = 2\sigma^2[1 - p + p^2 - (1-p)r(y,x_0) - pr(y,x_1) + p(1-p)r(x_0,x_1)]$$

The correlation coefficients not only relate the measured concentrations, but are also functions of the relative locations at which the concentrations are measured. Thus, we write

$$r(y,x_o) = r[ph]$$

$$r(y,x_1) = r[(1 - p)h]$$

$$r(x_o,x_1) = r[h]$$

where h is the appropriate scale factor, the distance between the points P_o and P_1. (p is a dimensionless parameter that is zero at P_o and 1 at P_1.) Then the mean square error of the interpolated value is $\overline{E^2}$. It is readily seen that at p = 0, r[ph] = 1, and $\overline{E^2}$ = 0 and that at p = 1, r[(1 - p)h] = 1 so that again $\overline{E^2} = 0$; that is, the interpolation error is zero at the two data points, which is as it should be.

The important feature of the above relation is that the mean square error of interpolation depends on the correlation coefficients as functions of the spacing between data points, $r[h]$, and of the distance of the interpolated point from the data points, $r[ph]$ and $r[(1-p)h]$.

Consider now a particular example of a correlation coefficient which is 1 at zero distance and reduces linearly to 0 at a distance ℓ, and we assume that ℓ is larger than h, the distance between P_o and P_1. Then

$$r[h] = 1-h/\ell$$
$$r[ph] = 1-ph/\ell$$
$$r[(1-p)h] = 1-(1-p)h/\ell$$

The mean square error of the interpolated value may then be written as

$$\overline{E^2} = 2\sigma^2 p(1-p)h/\ell$$

It is readily seen that the error of the interpolated value is a maximum at $p = \frac{1}{2}$ and has the value there of $\overline{E^2} = \sigma^2 h/2\ell$. Under these conditions, we have an explicit expression that can be used to determine the spacing of the observation points. Thus, if we specify the maximum allowable mean square error of interpolation, $\overline{E^2}$, then the distance between data points may not exceed the value $h = 2\ell(\overline{E^2}/\sigma^2)$.

The value of ℓ may be thought of as a "range of influence" of the correlation coefficient. The larger the value of ℓ, the farther apart the observation points may be spaced. The spacing also depends on the inherent variability of the data through the term σ^2. The more highly variable the data the closer the observation points to achieve the same maximum mean square error of interpolation.

Other analytical expressions may be used for the correlation coefficient and the location of the point of maximum error may be found. It is readily shown that the point of maximum mean square error is at $p = \frac{1}{2}$ and that the mean square error of interpolation will be given by

$$\overline{E^2} = 2\sigma^2 \left\{ 1 - (1/4)[1-r(h)] - r(h/2) \right\} .$$

This expression makes it possible to compute the spacing between observation points that must not be exceeded when a mean square error of interpolation is specified and the correlation coefficient function is known.

The more complicated interpolation formulas for a two dimensional array of points lead to vastly more complicated arithmetic, but do not change the essential ideas brought out by the above elementary analysis. The main idea is that the mean square error of interpolation depends on the structure of the correlation coefficient as a function of the distance separating the points at which the pollutant concentration is measured. When this structure is known, the spacing of the observation point to achieve a given mean square interpolation error may be specified.

B. Linear Regression As An Interpolation Formula

This section is devoted to an elementary derivation of the linear regression estimate of pollutant concentration at a point P based on observed pollutant concentrations at a network of points P_i, $i = 1, ---, n$. It is initially assumed that the values of pollutant concentration at P are observed. The point of view is then reversed and the regression equation is considered from the point of view of an interpolation formula which is used to estimate the pollutant concentration at P when it is not observed there.

1. Derivation of the Basic Relations

Let Y be the pollutant concentration at P and let X_i be the pollutant concentrations at points P_i, where the P_i are a network of n observation points, $i = 1, ---, n$. To simplify the situation we consider the standardized variables (departure from the mean divided by the standard deviation)

$$y = (Y-\bar{Y})/\sigma_Y \ , \ x_i = (X_i-\bar{X}_i)/\sigma_{X_i} \ , \ i = 1, ---, n$$

and we consider the relation

$$\hat{y} = b_1 x_1 + --- + b_n x_n \tag{1}$$

where $\hat{y}$ is the estimate of y given the values x_1, --- x_n.

The least squares procedure for determining the coefficients b_i leads to the set of equations

$$\begin{aligned} \overline{(yx_1)} &= b_1 \overline{(x_1^2)} + --- + b_n \overline{(x_n x_1)} \\ &--- \qquad\quad --- \\ \overline{(yx_n)} &= b_1 \overline{(x_1 x_n)} + --- + b_n \overline{(x_n^2)} \end{aligned} \tag{2}$$

where the bar over the symbol indicates a mean value and where y is assumed to be a measured value at P. Since the variables are normalized, these are correlation coefficients.

It will be convenient to write these equations also in the standard form

$$
\begin{aligned}
b_1 a_{11} + \text{---} + b_n a_{1n} &= g_1 \\
\text{---} \qquad \text{---} & \\
b_1 a_{n1} + \text{---} + b_n a_{nn} &= g_n
\end{aligned}
\tag{2a}
$$

and we note that the matrix of coefficients is symmetric, $a_{ij} = a_{ji}$.

The solution for the b_i's may be written in the long form using Cramer's rule

$$
b_i = \begin{vmatrix}
(\overline{x_1^2}), & (\overline{x_1 x_2}), & \text{---}, & (\overline{yx_1}), & \text{---}, & (\overline{x_1 x_n}) \\
(\overline{x_2 x_1}), & (\overline{x_2^2}), & \text{---}, & (\overline{yx_2}), & \text{---}, & (\overline{x_2 x_n}) \\
(x_n x_1), & \text{---} & & \text{---} & & \\
(\overline{x_n x_1}), & (\overline{x_n x_2}), & \text{---}, & (\overline{yx_n}), & \text{---}, & (\overline{x_n^2})
\end{vmatrix}
\div
\begin{vmatrix}
(\overline{x_1^2}), & \text{---}, & (\overline{x_1 x_n}) \\
\text{---} & & \text{---} \\
 & & \\
(\overline{x_n^2}), & \text{---}, & (\overline{x_n^2})
\end{vmatrix}
\tag{3}
$$

ith column.

and also in the form (Kenney and Keeping, 1951)

$$b_i = \sum_j g_j a^{ij} \tag{3a}$$

where $g_j = (\overline{y\, x_j})$, and the term a^{ij} is the element from the inverse of the matrix of coefficient (a_{ij}). This follows immediately from (3) by expanding the determinant in the numerator in terms of the sum of the products of the elements in the i'th column and the cofactors of this column. The ratio of the cofactor of the element of the i'th row of the j'th column to the value of the determinant yields the element a^{ij} of the inverse (Turnbull, 1960). The solution for finding $\hat{y}$ may then be written as

$$\hat{y} = \sum_i x_i \Big(\sum_j g_j a^{ij} \Big) = \sum_j g_j \Big(\sum_i x_i a^{ij} \Big) \tag{4}$$

The important point to be considered now is that the correlation coefficients $\overline{(y\ x_i)} = g_i$ (and also the correlation coefficients $\overline{(x_i x_j)} = a_{ij}$) are <u>functions</u> of the locations of the point P and the point P_i (or of the points P_i and P_j). Since the points P_i are fixed, we focus our attention on the point P and write

$$g_j = g_j(P,P_j)$$

Then (4) may be expressed as

$$\hat{y} = \sum_i x_i [\sum_j g_j(P,P_j)a^{ij}] = \sum_j g_j(P,P_j)[\sum_i x_i a^{ij}] \quad (4a)$$

Thus, (4a) may be considered as an <u>interpolation</u> formula since the location of the point P at which $\hat{y}$ is estimated appears explicitly in the correlation coefficient $g_j(P,P_j)$. If we know the functional form of the correlation coefficients as dependent on coordinates, then (4a) may be looked on as determining $\hat{y}$ from a linear combination of the observed concentration x_i with coefficients which depend on the location of P, or it may be considered as a linear combination of correlation coefficients, functions of the location P where the concentration $\hat{y}$ is estimated, weighted by factors that depend on the observed concentrations at the points P_i.

2. Interpretation of Results

The results of the simple linear regression for pollutant concentration $\hat{y}$ at a variable point P in terms of observed pollutant concentrations x_i at a fixed network of points P_i are discussed below. The point of view in this discussion is that the linear regression is a particular realization of a "Wiener filter" for the interpolation and smoothing of observed information on pollutant concentrations (Wiener, 1949). The elaborate mathematical apparatus of N. Wiener's original treatment is abandoned in favor of a more general point of view so that more general results are obtained (at least in a limited sense). We discuss several particular situations that illustrate the kind of results that can be obtained.

a) The point P at a point of P_i

If the point P at which concentrations y are measured coincides with a point of the observing network, say P_k, and if the data on y is identical with the values of x_k at P_k, it is readily seen that $\overline{yx_i} = \overline{x_k x_i}$ and that, from (3) $b_k = 1$, and $b_j = 0$ if $j \neq k$. (In the first case the kth column of the numerator in (3) is exactly that of the denominator; in the second the kth column of the numerator is exactly the same as the jth column and hence the determinant has the value zero.) In this case $y = x_k$ is the result of the use of the regression, which is precisely what it should be.

b) Continuous Correlation Coefficient

Consider the case in which the correlation coefficients $g_j(P,P_j)$ are continuous functions of P and for which $g_j \rightarrow 1$ for $P \rightarrow P_j$. A one dimensional schematic of this situation is shown in Fig. I-1a. The use of the interpolation

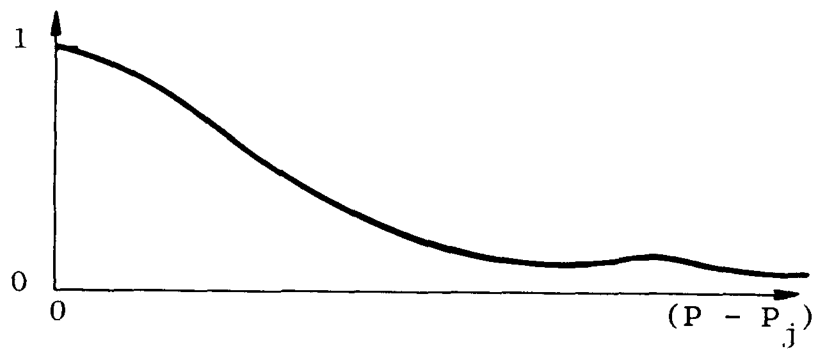


Fig. I-1a. Schematic illustration of the Correlation Coefficient $g_j(P,P_j)$ as a function of $(P-P_j)$.

formula (4) or (4a) leads to a smooth interpolation of the data at the points which lie between the data points P_i and the interpolated values lie on a surface that passes through the data values x_i. This is shown schematically in the Fig. 1-b for a one dimensional situation.

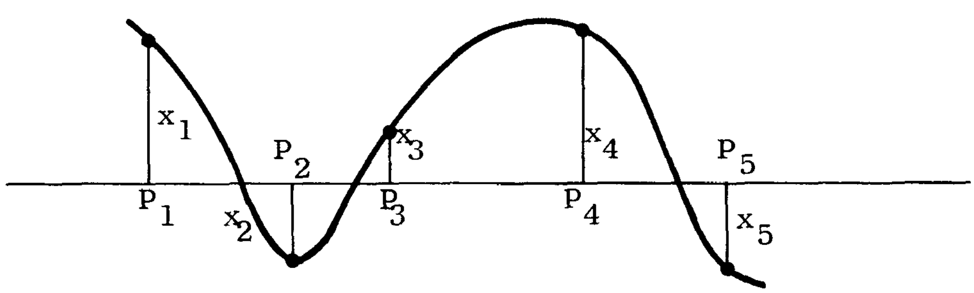


Fig. I-1b. Schematic illustration of the interpolation between data points for correlation coefficients that are continuous.

c) <u>Extremely Discontinuous Correlation Coefficient</u>

The correlation coefficient as a function of the location of P with respect to P_j as in $g_j(P,P_j)$ <u>must</u> be a continuous function of $P-P_j$, except that it may have a jump discontinuity at the origin $P-P_j = 0$. An extreme case is that in which $g_j(P,P_j) = 1$ when $P = P_j$ and is equal to zero if $P \neq P_j$. This is illustrated in Fig. I-2a. The resulting interpolation for

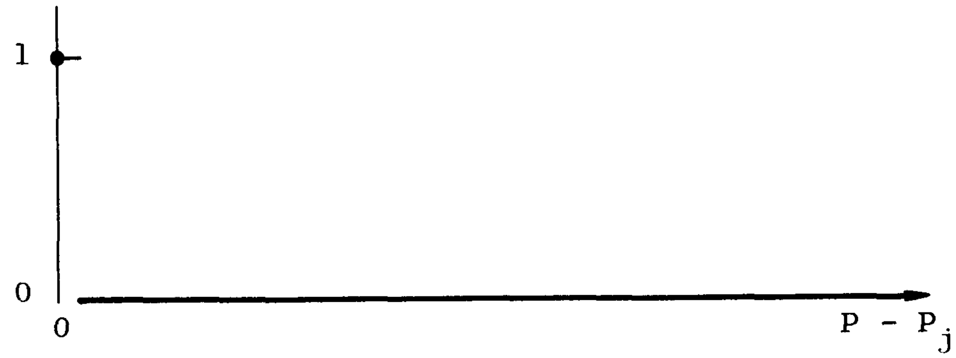


Fig. I-2a. Schematic illustration of the extreme discontinuity possible for a correlation coefficient in which it has the value 1 at $P-P_j=0$ and has the value 0 at $P-P_j \neq 0$.

this kind of a correlation coefficient is illustrated in Fig. I-2b. The interpolated values are zero between the data points, but at the data point the observed data values are obtained.

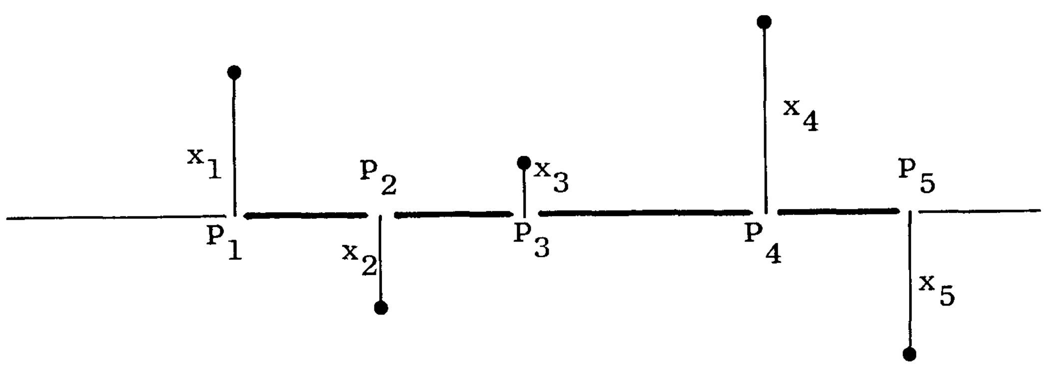


Fig. I-2b. Schematic of the interpolation between data points when the correlation coefficient is that illustrated in Fig. I-2a.

d) <u>Discontinuous Correlation Coefficient</u>

When measured values are subject to independent, random errors, the correlation coefficient has a (small) jump discontinuity at the $P = P_j$ which is dependent on the relative values of the standard deviation of the quantity being measured and the standard deviation of the random errors. Such a correlation coefficient is illustrated in Fig. I-3a.

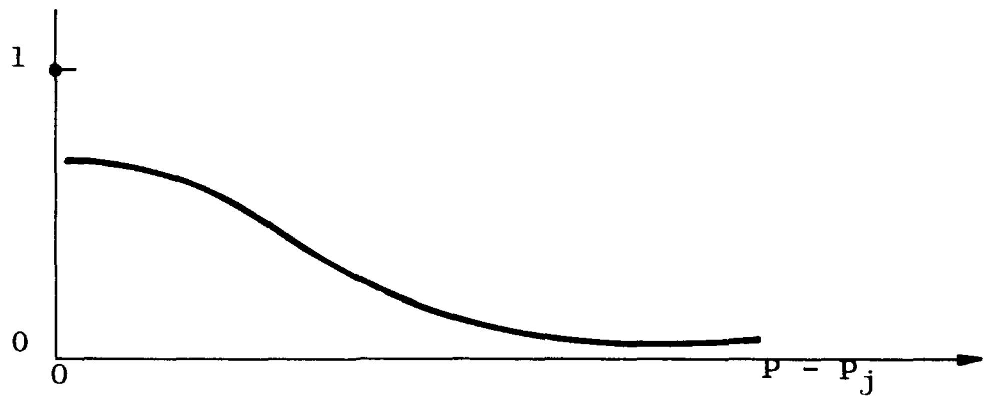


Fig. I-3a. Schematic illustration of a correlation coefficient function that arises when there are errors in measurement.

The use of such correlation coefficients in (4) or (4a) results in an interpolation which smooths the data between observation points as shown in Fig. I-3b.

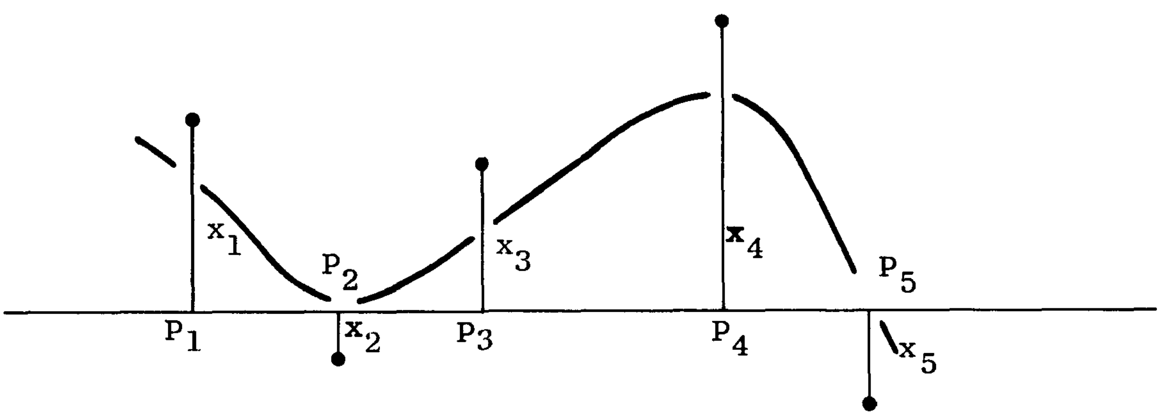


Fig. I-3b. Schematic showing the smoothing of interpolated values when the correlation coefficient has a small discontinuity as in Fig. I-3a.

e) Range of Influence

The "range of influence" of a correlation coefficient is of importance in the discussion of observation network density. This is loosely defined as the distance $P - P_j$ over which the correlation coefficient is significantly different from zero. A correlation coefficient with limited range of influence is illustrated in Fig. 4a. When the data points are spread out

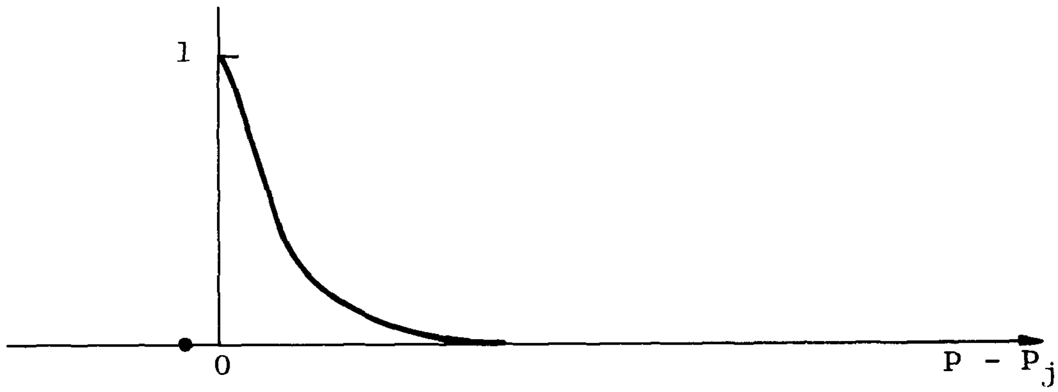


Fig. I-4a. Schematic of a correlation coefficient with a limited "range of influence".

so that no one is within the "range of influence" of another, the interpolation formula (4) and (4a) leads to results illustrated in Fig. I-4b. It is immediately apparent that for an adequate

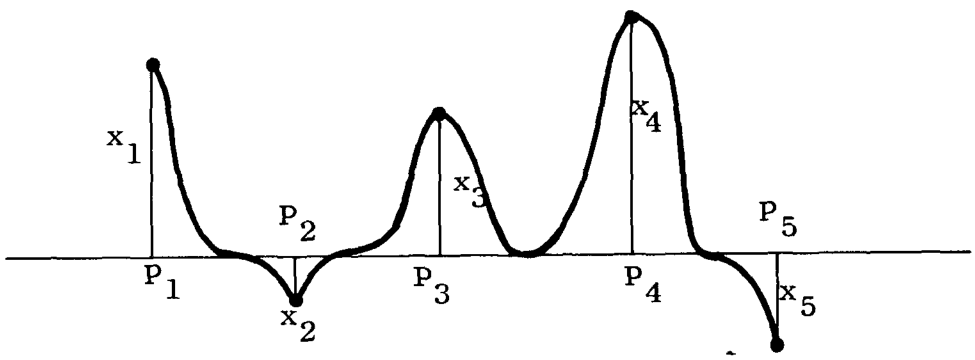


Fig. I-4b. Schematic of interpolation when data points are so sparsely located that one is not within the "range of influence" of another.

network of observation points, the distance between points must be small enough that one or more data points must be within the "range of influence" of some other data point.

f) Small Scale Effects

In dealing with atmospheric problems, the influence of small scale effects must be adequately accounted for (or adequately smoothed out). These show up as a small hump on the correlation coefficient peaking it sharply upward at $P - P_j = 0$ as illustrated in Fig. I-5a.

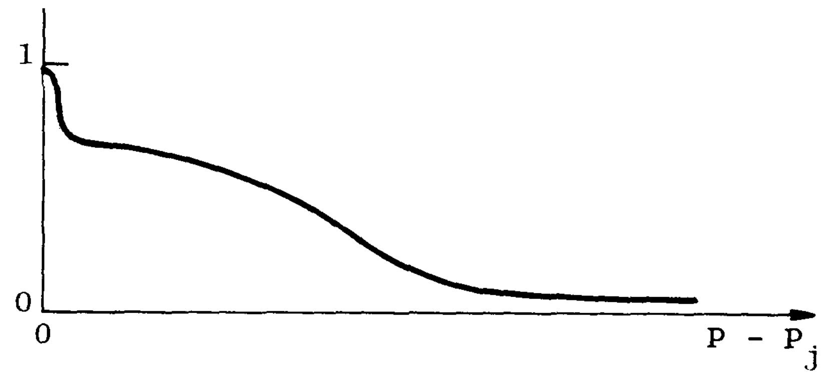


Fig. I-5a. Schematic illustration of a correlation coefficient showing both large and small scale effects.

The effect on the interpolation formula (4) or (4a) is shown in Fig. I-5b.

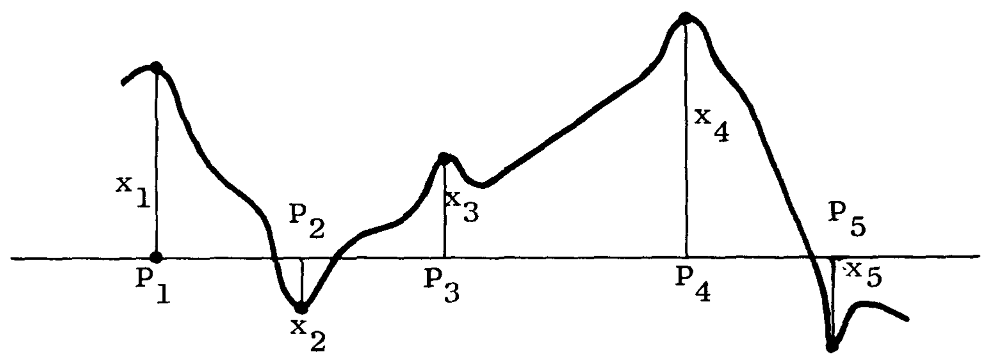


Fig. I-5b. Schematic illustration showing the results of small scale effects on the interpolation of data.

g) Summary

The implications of the above illustrative examples for the determination of observation network density are immediate. The first result is that the network of observation points should be sufficiently dense that each data point is within the "range of influence" of another data point. The second important result is that the effects of errors and small scale effects must be carefully determined and the interpolation formula used in such a way that there is adequate filtering of these effects in interpolating the results of observations.

3. Error of Interpolated Values

The mean square error of the interpolated values may be written at once from the results of the least squares estimation (interpolation) formulas (1), (2), (2a), (3), and (3a). The expression for the mean square error is

$$\overline{E^2} = \overline{y^2} - (b_1 g_1 + --- + b_n g_n) \tag{5}$$

$$\overline{E^2} = \overline{y^2} - \sum_i b_i g_i \tag{5a}$$

and using (3a) for the value of the b_i's, one obtains

$$\overline{E^2} = \overline{y^2} - \sum_i \sum_j g_i a^{ij} g_j \tag{5b}$$

The expression (5b) gives the mean square error $\overline{E^2}$ in terms of the mean square deviation of pollutant concentration at the point P, $\overline{y^2}$, the correlation coefficient involving concentration at the point P and the measurement points P_i, $g_i(P,P_i)$, and constants involving the geometry of the measurement points a^{ij}. The value of the mean square deviation, $\overline{y^2}$, is readily estimated from the field of observation point values $\overline{x_i^2}$. Since we have used the normalized form, the values of $\overline{x_i^2}$ are all 1 and the value of $\overline{y^2}$ would be taken as 1 also.

The expression (5b) may be written to show the dependence of $\overline{E^2}$ on the location of the point P, by displaying this dependence in the correlation coefficients g_i, g_j

$$\overline{E^2}(P) = \overline{y^2} - \sum_i \sum_j g_i(P,P_i) a^{ij} g_j(P,P_j) \tag{5c}$$

The mean square error of interpolation for pollutant concentration is illustrated schematically for a one-dimensional example in Fig. I-6 (assuming a continuous correlation coefficient function as in Fib. I-1b or Fig. I-5b).

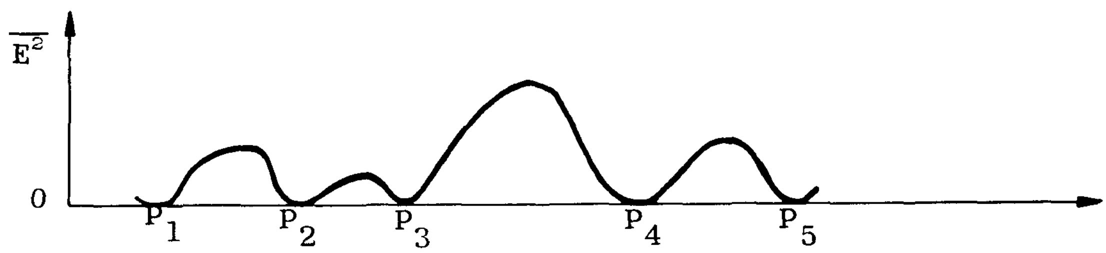


Fig. I-6. Schematic illustration of the mean square error of interpolated values for a one-dimensional case.

a) Implications for Measurement Locations

The implications of the above for determining the location of additional measurement points is reasonably obvious. We assume that it is desired to find the location of a few more observation points to improve the observation network. On the basis of (5c) one may easily compute the field of values of $\overline{E^2}$ as a function of the location of an interpolation point P. The point where $\overline{E^2}(P)$ is a maximum is the point where there is the greatest error in interpolated values. This could then be a point where an observation of pollutant concentrations could contribute the most information to the augmented network systems.

There are many practical considerations that must be taken into account in the location of observation points. These may be readily accounted for in the computation procedures. One may, for example, prescribe in advance the location of many "feasible" observation points and compute the mean square errors of interpolation at these "feasible" points. The "feasible" points at which the mean square error of interpolation is largest would then be candidates for the augmented set of additional observation points.

Note that the selection of additional observation points is a step-wise procedure. One locates the first additional observation point where the mean square error of interpolation is maximum (a "feasible" point). This point is then added to the network so that there are now n+1 points considered and the computation of $\overline{E^2}$ is redone on the basis of the augmented network. The maximum of $\overline{E^2}$ is found for this augmented case and a second (feasible) observation point located. The network now contains n+2 points and $\overline{E^2}$ is computed on this basis, etc., etc.

(Note: This kind of iterative procedure can be shown to lead to a solution that is not necessarily optimum in the general case, but it is a very practical approach and the results are usually not far from an optimum.)

CHAPTER II

DETAILED FORMULATION OF THE PROBLEM

The problem of determining an optimum network of air pollution observation sites consists, first of all, of determining what particular characteristics of the pollution field are of primary interest, how the data are to be used to obtain an estimate of these characteristics, and then to specify what it is that is to be optimized and with what restrictions this optimum is to be obtained. An extensive monograph might be written in the exploration of the various aspects of the problems involved in the items that have been enumerated above. Rather than go into all of this detail, we will short-cut these considerations and set down some rather simplified ground rules that will be followed.

A. Area Covered

In order to formulate the problem of specifying the optimum location of air pollution observation stations it is necessary to start by defining the specific area over which the pollution is to be observed. This is sufficiently obvious that it scarcely needs any explanation. If one is to optimize the location of air pollution stations in the area of Dirty City, they need to be located in the vicinity of Dirty City and not in Clean County, because Clean County is not the area with which we are concerned. In fact, one must be even more specific and define exactly what constitutes "in the area of Dirty City". This involves specifying a size, shape, and location. It may be, for example, that Dirty City is a compact area of more or less uniform diameter in any direction with this area centered approximately at the City Hall. In such a case, an adequate specification might well consist of the statement that the area to be covered is a circular area centered at the Dirty City City Hall with a radius of 23 km (say). This implies that one is to optimize the location of air pollution observation stations throughout this particular circular area. It may well be that one is also interested in air pollution measurements at distances farther than 23 km from City Hall, but these locations will not be a part of this particular air pollution station location problem.

B. Observation Collection

It is assumed that the data collected from the air pollution measuring sites will be measured simultaneously at all locations. The term simultaneous is used somewhat loosely and depends to some extent on the nature of the measurements to be used. If, for example, one is concerned with one-hour integrated air pollutant concentrations, the averaging should be done over the same clock hour at all sites; but a difference of five minutes or so between stations for the beginning and ending of this one hour integration period is relatively unimportant. In other words, the differences in time of beginning and ending of the totalizing interval are a relatively small fraction of this interval itself. If the averaging is considered on a 24 hour basis, then the exact beginning and ending of the averaging interval at each location can be correspondingly relaxed. On the other hand, if observation data at the different locations are taken at different clock hours, then it may well be that a comparison between stations is meaningless.

C. Source Locations, Types, Rates

It is assumed that the locations, types and emission rates of the pollution sources are known or at least can be estimated with reasonable accuracy. The basic reason for this is the fact that it is completely impossible to even approach the problem of the intelligent location of pollution observation sites without this kind of information. If one were interested in pollution at only one point, then obviously one locates a measuring site at this point. When the problem is to locate a network of measuring points, then the question immediately arises concerning the relation of pollution measurements at the several points with each other. If two (or more) pollution measurements essentially duplicate each other, then one is unnecessary (or at least of much less value). On the other hand, if measurements at two locations are completely unrelated to each other, then obviously additional measurements are desirable between them. The locations, types and emission rates of the sources involved are quantities which determine how the measurements at related points are going to be related, or unrelated, to each other. A more exact specification of the relationship between the pollution measurements at different locations is discussed at some length in Section G below.

D. Meteorological Conditions

The meteorological conditions that prevail over the area concerned are responsible for determining how the pollutants are carried from their source locations to other points and whether they tend to be concentrated near the ground or carried high into the atmosphere. The principal factors concerned here are the winds in the lower atmosphere and the stability (or unstability) of the lower layers of the atmosphere. These conditions need to be carefully specified and should adequately describe the conditions over the area concerned. If in the area concerned the wind has a strongly predominant direction, it would seem reasonable to locate pollution observation points down-wind from the more important pollution sources. Since stable conditions tend to hold pollutants near the ground, it would seem reasonable to locate observation points in the down-wind direction from the important sources where this wind direction is that which prevails when the air is stable. Consequently, the meteorological conditions need to be specified as the frequency of occurrence of various classes dependent on wind direction, wind speed, and stability.

E. Type of Pollution Measurements

The type of pollution measurements to be made is another factor contributing to the optimum location of pollution measuring locations. If hourly average concentrations are of primary interest, it would appear necessary to have more closely located observation points than if, say, 24 hour average concentrations are considered the more important.

F. Interpolation/Extrapolation Methods

The method that is used to interpolate data between observation points is critical for the formulation of the optimization technique for site locations. There is an almost limitless number of techniques that may be used. The method of interpolation used in this study consists of using a linear estimate of the pollution concentration based on the concentrations at the points where the concentrations are observed and for which the correlation coefficient function or covariance function is known (or at least can be approximated with reasonable accuracy). Various aspects of this process were discussed in Chapter I with illustrations in the case of a one-dimensional field (to simplify the figures). This provides a highly flexible procedure that can be adapted to a wide variety of different situations.

The interpolation/extrapolation method then consists of estimating a measure of pollutant concentration at a point (ξ,η) in the region to be covered, say $\hat{Y}=Y(\xi,\eta)$, from a linear combination of observations, $X_i = X(\xi_i,\eta_i)$ made at observation points P_i with coordinates (ξ_i,η_i). The formula for this estimate is

$$\hat{Y} = b_o + b_1X_1 = --- + b_nX_n \qquad (1)$$

(where it assumed that there are n observation points). The coefficients b_i, $i=0,1, ---, n$, will be determined later. The numbers that describe the pollution concentration vary over several orders of magnitude and are by no means normally distributed, the logarithm of the pollutant concentration is actually used as the measure of concentration. The logarithm of concentrations is more nearly normally distributed (i.e., the pollution concentrations tend to be distributed in a log-normal way).

It is assumed that reasonably accurate estimates of the mean concentration measure are known. These will be denoted by $\bar{Y}$, $\bar{X}_1$, ---, $\bar{X}_n$ so that one may use the departures from the mean, $\hat{y} = \hat{Y}-\hat{\bar{Y}}$, $x_1 = X_1-\bar{X}_1$, ---, $x_n = X_n-\bar{X}_n$ with the result that

$$\hat{y} = b_1x_1 + --- + b_nx_n \qquad (2)$$

The coefficient b_o that appeared in (1) is determined from

$$\bar{\hat{Y}} = b_o + b_1\bar{X}_1 + --- + b_n\bar{X}_n \qquad (3)$$

after the coefficients b_1, ---, b_n have been determined. (Actually, the problem at hand does not require that any of the coefficients be computed explicitly; the technique used is more clearly described if they are included at this point.) If one had an observation point at the point P, coordinates (ξ,η), where $\hat{Y}$ (or $\hat{y}$) is being estimated, the observed value would be Y (or y). If we assume that actual values y, x_1, ---, x_n are observed in an ensemble of situations that have been defined by the conditions of Sections C and D above, then the coefficients b_1, ---, b_n would be determined from the normal equations

$$\begin{aligned} (\overline{x_1 y}) &= b_1(\overline{x_1 x_1}) + --- + b_n(\overline{x_1 x_n}) \\ &--- \qquad\qquad\qquad\qquad (4) \\ (\overline{x_n y_n}) &= b_1(\overline{x_n x_1}) + --- + b_n(\overline{x_n x_n}) \end{aligned}$$

where $(\overline{x_i x_j})$ is the covariance of the concentration measure at the points P_i and P_j and $(\overline{x_i y})$ is that for the points P_i and P. The normal equations (4) are obtained by requiring that the square of difference between the concentration measure and the estimate from equation (2) summed over the ensemble of situations, be a minimum in terms of the coefficients b_1, ---, b_n. (See standard texts on this subject, for example Kenney and Keeping, 1951.)

If we let $a_{ij} = (\overline{x_i x_j})$ and $g_i = (\overline{x_i y})$, then equations (4) may be written as

$$g_i = \sum_j a_{ij} b_j, \qquad i,j=1, ---, n \qquad (5)$$

These equations may be solved for the coefficients b_1, ---, b_n, the formal solution being expressed as

$$b_i = \sum_j a^{ij} g_j, \qquad i,j=1, ---, n \qquad (6)$$

The mean square error of estimate is given by the expression

$$\overline{e^2} = \overline{(y-\hat{y})^2}$$

where $\hat{y}$ is the expression (2). This may be expressed in the form

$$\overline{e^2} = \overline{(y^2)} - \sum_i b_i g_i \qquad (7)$$

(Kenney and Keeping, 1951, or other suitable text). When the expression (6) for the b_i's is substituted into (7), the result is

$$\overline{e^2} = \overline{(y^2)} - \sum_i \sum_j g_i a^{ij} g_j. \qquad (8)$$

In (7) and (8) the term $\overline{(y^2)}$ is the variance of the pollution concentration measure at the point P.

(<u>Digression</u>. The quantities y, x_i above have been expressed in terms of departures from the mean as per the expressions in the text between equations (1) and (2).

The quantities $\overline{(y^2)}$ and $\overline{(x_i x_i)}$ are the variances of the concentration measures at the various points concerned. In terms of standard deviations, σ and σ_i, $\overline{y^2}=\sigma^2$, $\overline{(x_i x_i)}=\sigma_i{}^2$. These may be used to convert the concentration measures to standardized form $(Y-\overline{Y})/\sigma$, $(X_i-\overline{X}_i)/\sigma_i$, and the argument remains unchanged with the exception that in (7) and (8) the value 1 is substituted for $\overline{(y^2)}$ and the error of estimate $\overline{e^2}$ becomes $\overline{e^2}/\sigma^2$.)

The role of the coordinates of the points concerned in the preceding relations is important, but is concealed in the notation. The covariances that appear in the normal equations (4) and subsequently are functions of the locations of the points for which the index number appears as a subscript with the exception of the point P for which no subscript appears. Thus $a_{ij} = \overline{x_i x_j} = a_{ij}(P_i,P_j)$ or $= a_{ij}(\xi_i,\eta_i;\xi_j,\eta_j)$ while $g_i = \overline{x_i y} = g_i(P_i,P)$ or $= g_i(\xi_i,\eta_i;\xi,\eta)$. The elements of the inverse matrix, a^{ij}, are exceptions to this. In the process of matrix inversion, all of the elements of the matrix are involved so that each element of the inverse, a^{ij}, involves all of the points P_1, ---, P_n, but not the point P at which the concentration measure is being estimated. This point appears only in the terms $g_i = g_i(P_i,P)$. If one substitutes the expression (6) for b_i into the equation for the estimated concentration measure, (2), the result is

$$\begin{aligned}\hat{y} &= \sum_i x_i \Big(\sum_j a^{ij} g_j \Big) \\ &= \sum_i x_i \Big(\sum_j a^{ij} g_j (P_j,P) \Big)\end{aligned} \qquad (9)$$

where in the second line the location of the point P at which $\hat{y}$ is estimated is explicitly shown. If the covariance function $g_i(P_j,P)$, as a function of the points P_i and P, is known or can be estimated reasonably well, it may be interpreted as an interpolation formula for $\hat{y}$ based on a linear combination of the observed concentration measures, x_i, at the points P_i. The expression in parenthesis gives the "weight" of each observed measure, x_i, in terms of the location of the point P with respect to the other points, P_i, i=1, ---, n. This is quite analogous to the standard two-dimensional interpolation formulas as for example those shown in Section A of Chapter I.

One may also note that if in (9) the point P is the same as the point P_k, then $g_j(P_j,P)$ becomes $g_j(P_j,P_k)$. Now by virtue of its definition [preceding equation (5)] this is only another way of writing a_{jk}, that is

$$g_j(P_j,P_k) = \overline{(x_j x_k)} = a_{jk}.$$

The summation in parenthesis in equation (9) then reduces to

$$\sum a^{ij} a_{jk} = \delta_{ik}$$

where $\delta_{ik} = 0$ if $i \neq k$ and $= 1$ if $i = k$. This means that the expression (9), when summed on index i, ignores all of the observed values at other points and assigns to the point P ($=P_k$) the estimate $\hat{y}=x_k$, i.e., the value observed. Note, however, that the covariance function $g_k(P_k,P)$ is not necessarily continuous at the point P_k, so that as P approaches P_k the value $\hat{y}$ need not approach x_k. The details of this situation were discussed in Chapter I.

G. Optimization Method

The expression for the mean square error of estimate given by equation (8) forms the basis of the optimization method that will be used. The development of this relation was based on the assumption that the covariance of the concentration measurements between any two points, P_i and P_j, was known, $a_{ij} = \overline{x_i x_j}$, so that the elements of the inverse matrix, a^{ij}, could be found. It was also assumed that the covariance function for concentration measure between an observation point, P_i, and any other point, P, in the area concerned was also known or at least could be reasonably well estimated. (The details of how these assumptions are realized are discussed in Chapter III.) The equation (8) then permits one to compute the error of estimate using (9) as an interpolation formula for any arbitrary location of the point P.

One may say that the selection of the n observation points P_1, ---, P_n has been chosen in an optimal manner if the largest mean square error of estimate, $\overline{e^2}$, at any point P not an observation point has been reduced to a satisfactory level. There are many ways in which such an optimization procedure can be carried out. The one adopted here is a one-step-at-a-time add-on method. It is presumed to start with that there are a few existing observation points or at least a few points at which observations will be made based on prior considerations. On the basis of this given starting network of observation points, the equation (8) is used to locate the point at which the error of estimate is largest. This point is then accepted as a best location for a new observation point. This point is then added to the list of observation points that are used to determine the error of estimate from equation (8). The process is then repeated; the location of the point of maximum mean square error is found, a new observation point is located, etc. The iterations of this process are terminated when the largest mean square error of estimate has been reduced to an acceptable level.

CHAPTER III

THE CORRELATION OF POLLUTANT CONCENTRATION BETWEEN POINTS

In order to implement equation (8) of Chapter II for use in estimating the mean square estimate, it is necessary to know the correlation coefficient for pollutant concentration measures between observation points and between an observation point and an arbitrary point in the area of interest. The use of actual data from past measurements of pollution concentrations is desirable, but, in dealing with an area over which observations have not been made, it is not possible to follow such a procedure. Consequently, the correlation coefficients for pollution concentrations were determined by using a model to generate synthetic data and the correlation coefficients were computed from the synthetic data.

A. The Model Used to Generate Synthetic Data

The model that was used to generate the synthetic data on which the correlation coefficients were based was a simple Gaussian Plume model described by equation (3.2), page 6, of Turner (1970) as follows

$$\chi(x,y,0,H) = (Q/\pi\sigma_y\sigma_z u)\exp[1/2\{(y/\sigma_y)^2+(H/\sigma_z)^2] \qquad (1)$$

where (x,y) are coordinates of the point at which the concentration $\chi(x,y,0,H)$ is calculated (x is measured down-wind from the source, y is measured cross-wind from the down-wind axis), Q is the source strength, H is the source (stack) height, u is the wind speed, σ_y, σ_z are dispersion coefficients. The dispersion coefficients, σ_y, σ_z, were computed from the formulas developed by Eimutis and Konicek (1972).

The wind conditions were computed on the basis of a double circularly normal density function. This is expressed as

$$f(u,\theta)udud\theta = [k_1 f_1(u,\theta;w_1,\phi_1,\sigma_1)+k_2 f_2(u,\theta;w_2,\phi_2,\sigma_2)]udud\theta$$

where k_1,k_2 represent the proportion of the time the wind is in the state described by $f_1(u,\theta;---)$ and $f_2(u,\theta;---)$ respectively and $k_1+k_2=1$. The functions $f_1(u,\theta;---)$, $f_2(u,\theta;---)$, are the circularly normal bivariate density functions with parameters as shown after the semicolon, but expressed in terms of wind speed, u, and direction, θ, instead of in terms of rectangular wind components. Thus

$$f(u,\theta;w,\phi,\sigma)=(\pi\sigma^2)^{-1}\exp\{-\sigma^{-2}[u^2+w^2-2uw\cos(\theta-\phi)]\}$$

where u is the wind speed, θ is the wind direction, w is the mean resultant wind speed, ϕ is the mean resultant wind direction, and σ is the vector standard deviation. (See Brooks, et al, 1946, or Brooks, et al, 1950.) A single circularly normal density function for wind at St. Louis is inadequate since there are two distinct modes of the over-all wind density. One of these is at $w_1=7.90$, $\phi_1=198.5°$, $\sigma_1=6.5$, $k_1=0.4$ while the other is at $w_2=9.00$, $\phi_2=325°$, $\sigma_2=6.50$, $k_2=0.6$. Stability class 4 was used throughout (Turner, 1970).

The use of the probability density function for wind enables one to use far fewer parameters to describe the wind direction and velocity frequency tables. For 16 direction categories and 9 speed categories, a total of 144 entries are required. In this case of a bimodal density function, only 8 parameters are needed while the speed and direction categories may be divided into arbitrarily small intervals.

There were 21 pollutant sources used to compute the correlation coefficients. Their locations, strength and stack height are listed in Table III-1. Correlation coefficients between points were computed for a 9X9 point grid of locations centered at the "arch" in St. Louis. The spacing between points was 10 km. Thus an 81X81 matrix of correlation coefficients was obtained. (See Section D below for further detail of the program.)

Since the logarithm of pollutant concentration rather than concentration itself was to be used in computing the correlation coefficients, a very small background pollutant concentration was added in each case to avoid the difficulty presented by taking the logarithm of zero.

B. Examples of the Correlation Coefficients

Some of the correlation coefficients obtained in the way described in Section A are illustrated in Figures III-1 and III-2. The points of the grid were numbered serially starting in the lower left corner with point no. 1 and numbering upward in each column. Thus the bottom row of points are those numbered 1, 10, 19, 28, 37, 46, 55, 64, 73. The top right corner is point 81. The arch is at location 41.

Actual correlation coefficients for observed 24 hr SO_2 concentrations were also computed using information from the St. Louis 1964-1965 sulphur dioxide field data (Ruff, 1973a).

TABLE III-1

SOURCE STRENGTH PARAMETERS

POINT	STRENGTH	XS	YS	HS
1	.3806E+04	89.	168.	183.
2	.1930E+03	100.	164.	48.
3	.2500E+02	93.	154.	71.
4	.7300E+02	89.	148.	65.
5	.6360E+03	95.	147.	68.
6	.4630E+03	95.	145.	43.
7	.3650E+03	94.	145.	71.
8	.1730E+03	93.	144.	9.
9	.2540E+03	90.	140.	21.
10	.5595E+04	87.	130.	108.
11	.2433E+04	84.	120.	109.
12	.4900E+02	86.	114.	100.
13	.5900E+02	75.	150.	84.
14	.1200E+03	106.	159.	10.
15	.1200E+03	76.	149.	10.
16	.1200E+04	65.	149.	10.
17	.1200E+04	95.	149.	10.
18	.1200E+03	105.	149.	10.
19	.1200E+03	86.	139.	10.
20	.1200E+03	96.	139.	10.
21	.1200E+03	115.	129.	10.

MINIMUM CONCENTRATION = 1.000E-20

(1) The data were provided by EPA from the 1964-1965 St. Louis Air Pollution Study.

(2) Source strength and meteorology were assumed to be independent.

(3) XS, YS are the coordinates of the source locations in grid units. HS is the source height in meters. Source strength is in units of grams per second.

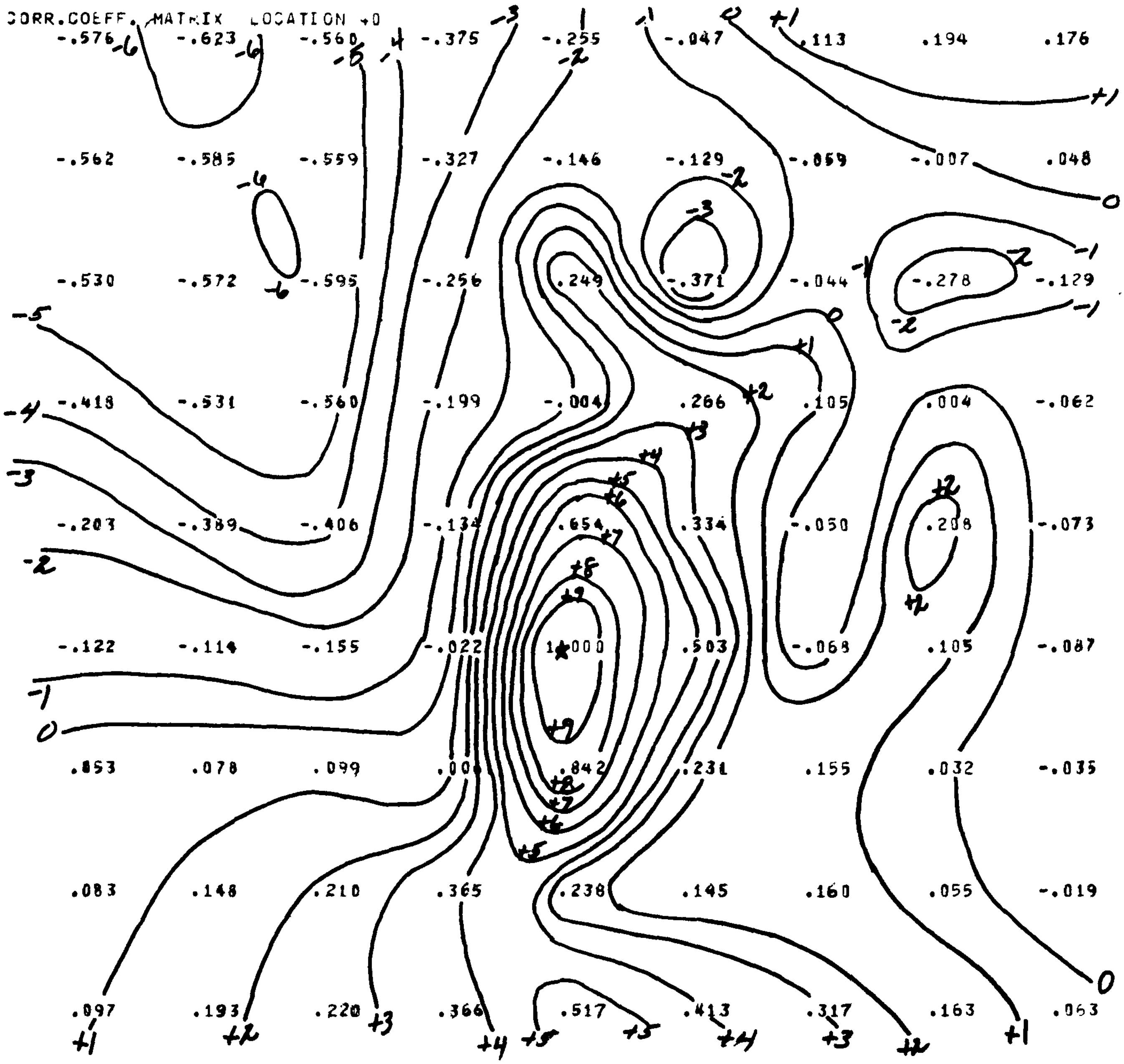


FIGURE III-1. Unsmoothed Contours of Correlation Coefficient Centered at Location 40

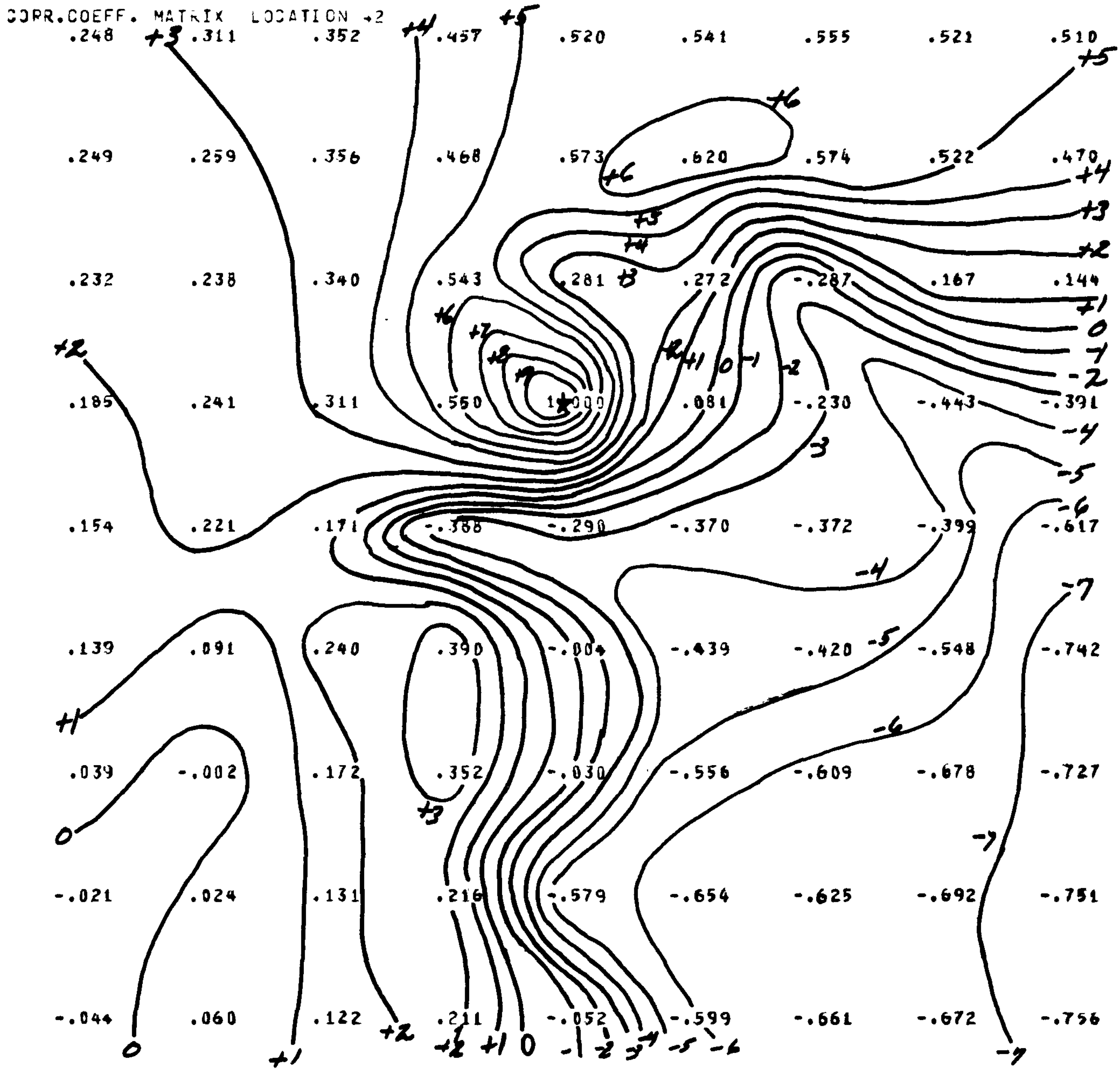


FIGURE III-2. Unsmoothed Contours of Correlation Coefficient Centered at Location 42

The correlation coefficient contours from actual SO_2 measurements are shown in Figures III- 3 and III-4 for locations No. 7 and 12 respectively. Location No. 7 has coordinates X=18.2, Y=22.50 and is described as "Power pole with transformer in front of 1914 Obear St., between 20th and Blair". Location No. 12 has coordinates X=22.48, Y=18.64 and is described as "Pole South of East Side Health District, 628 N. 20th St., East St. Louis, supplies power to building approximately 100' east of street". The scale is shown on these figures in the lower right hand corner since it is quite different from that of Figures III-1 and III-2. On these charts, the entire area corresponds to an area approximately 3 squares wide and 2 1/2 squares high on the preceding charts. The station numbers and locations for Figures III-3 and III-4 are shown in Figure IV-5, p. 92.

Figures III-1 and III-3 together with Figures III-2 and III-4 illustrate the large changes that occur in the correlation coefficient contours when the location of the point with which all other locations are being correlated crosses the central part of the area in which the pollution sources lie.

C. Interpolation of the Correlation Coefficients

The use of equation (8), Chapter II, to obtain the mean square error of estimate of pollutant concentration requires that the correlation coefficients be those between selected observation points or between observation points and an arbitrary point. The correlation coefficients (synthetic) calculated are those for points on a 9X9 grid with 10 km separation between rows/columns of points. To go from the latter to the former requires that some kind of

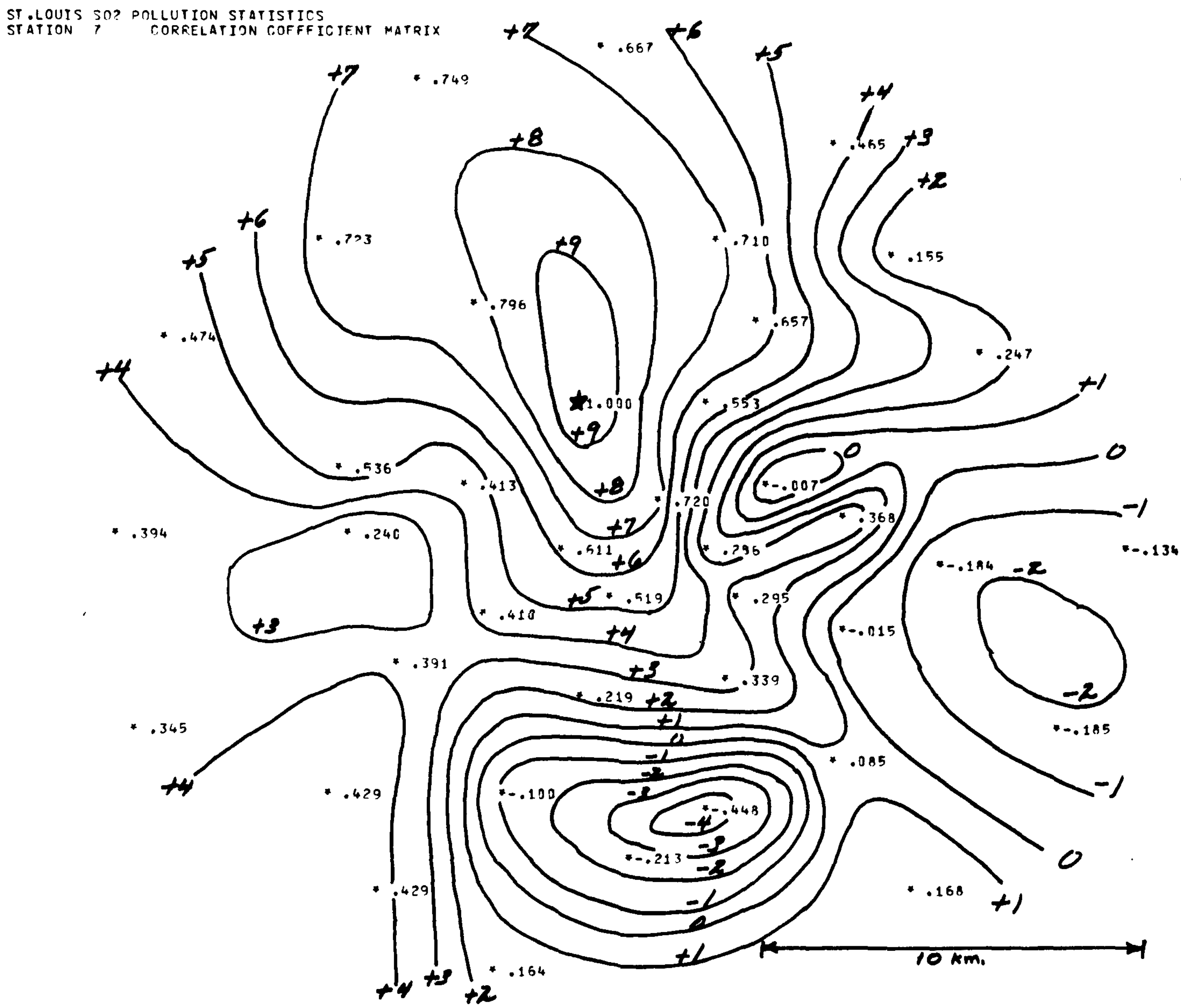


FIGURE III-3. Observed Correlation Coefficient of 24 hr. SO_2 with that at Station No. 7. Similar to the Simulated Correlation Coefficients in Figure III-1.

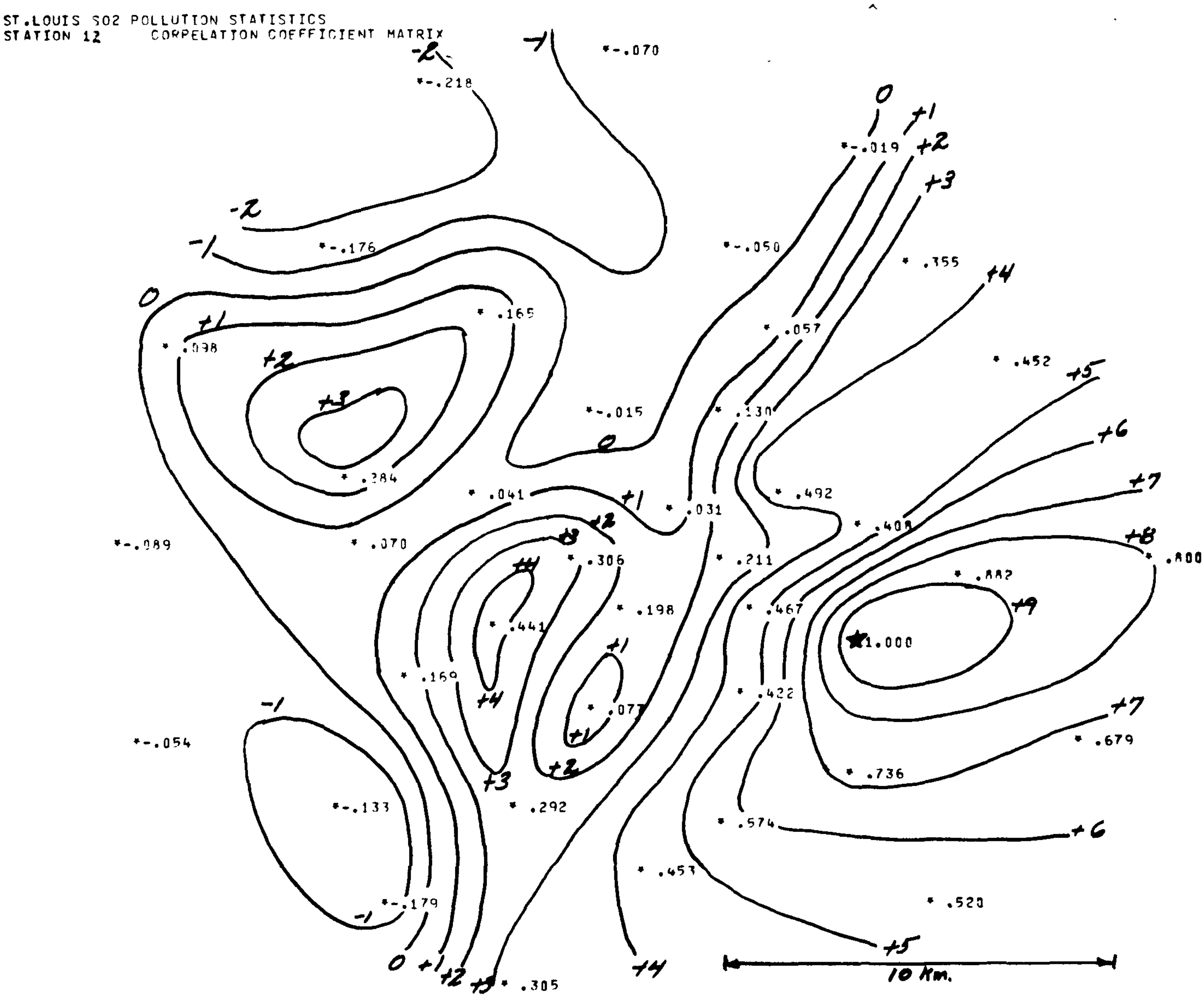


FIGURE III-4. Observed Correlation Coefficient of 24 hr. SO_2 with that at Station No. 12. Similar to the Simulated Correlation Coefficients in Figure III-2 interpolation procedure be used. The method finally adopted is based on the proper function/proper value representation of the correlation coefficient matrix.

To illustrate what is involved in the interpolation process, it is desired to obtain the correlation coefficient $r(x_i,y_i;x_j,y_j)$ where (x_i,y_i) are the coordinates of one observation site, P_i, and (x_j,y_j) are the coordinates of another observation site, P_j. We have available correlation coefficients between points on a 9X9 grid. Let these points have coordinates (ξ_m,η_m), (ξ_n,η_n) so that the correlation coefficients are functions $r(\xi_m,\eta_m;\xi_n,\eta_n)$. One is then faced with an interpolation procedure that involves four separate coordinates. This in itself is a reasonably formidable task. In view of the great irregularity of the correlation field, as illustrated in Figures III-1 and III-2, the task is even further complicated.

The proper function/proper value representation of the correlation coefficient matrix reduces the problem to that of performing two interpolations, each in a two-dimensional field, in succession. It also has the advantage that a certain amount of smoothing of the correlation coefficient field is being done at the same time. The details of this process are discussed in Chapter IV, so that only a brief extract of the essentials is given here. The essence of the situation is that the correlation coefficient matrix with elements $c_{mn}=\overline{x_m x_n}=r(\xi_m,\eta_m;\xi_n,\eta_n)$ may be represented in the form

$$c_{mn}=r(\xi_m,\eta_m;\xi_n,\eta_n) = \sum_k \lambda_k \phi_k(\xi_m,\eta_m)\phi_k(\xi_n,\eta_n) \qquad (2)$$

where, on the right hand side the summation over the index k involves the proper values λ_1,λ_2, ---, λ_k, ---, λ_p, all positive numbers and in decreasing order, $\lambda_1 > \lambda_2 ---> \lambda_p$, and the proper functions $\phi_k(\xi_m, \eta_m)$, one for each of the proper values. (These are also referred to as eigenfunctions and eigenvalues, principal values and principal components, empirical orthogonal functions, etc.) These are computed at the points of the 9X9 grid over which the matrix of synthetic correlation coefficients was obtained (81x81). Each of these is dependent on only the two coordinates of a grid point. Then to obtain the synthetic correlation coefficient between observation sites $a_{ij}=r(x_i,y_i;x_j,y_j)$ one simply interpolates among the grid points to obtain each $\phi_k(x_i,y_i)$ and each $\phi_k(x_j,y_j)$ and substitutes back in equation (2). To obtain the factors g_i, g_j that appear in (8), Chapter II, one notes that these are also correlation coefficients, but involve an observation site (x_i,y_i) and an arbitrary point (x,y). The coordinates of the arbitrary point are simply used to get an interpolated value of $\phi_k(x,y)$ for each k.

The summation in equation (2) does not necessarily run from 1 to 81. The upper value, p, is much smaller than 81 and is discussed in detail in Chapter IV.

In view of the fact that even the proper functions are somewhat irregular, quadratic interpolation based on six points was used. The formula (25.2.67) from Abramowitz and Stegun (1964), p. 882, was used:

$$f(x_o+ph,y_o+qh)=[q(q-1)/2]f(0,1)+[p(p-1)/2]f(-1,0)$$

$$+(1+pq-p^2-q^2)f(0,0)+[p(p-2q+1)/2]f(1,0)$$

$$+[q(q-2p+1)/2]f(0,1)+pqf(1,1).$$

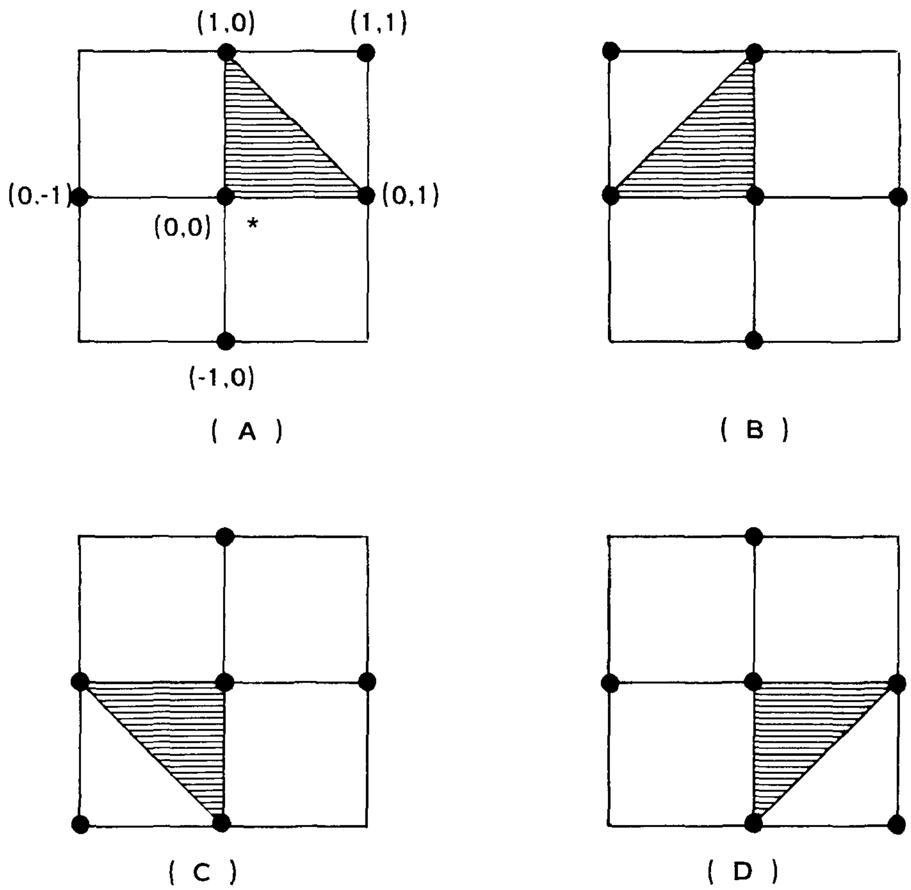


FIGURE III-5

THE SIX POINT ARRAY FOR QUADRATIC INTERPOLATION (A) AND THE THREE 90° ROTATIONS ABOUT (0,0). THE AREA OF BEST INTERPOLATION IS INDICATED IN EACH.

The arrangement of the points is shown in (a) of Figure III-5. This formula gives the best results over the triangle bounded by (0,0), (1,0), and (0,1) (shaded in the figure). In order to maintain this degree of interpolation accuracy, corresponding formulae were written for the three 90° rotations of this unsymmetrical array of points as shown in (b), (c), and (d) of the figure. To carry out an interpolation, the point (0,0) on the grid was located so that (x,y) fell within or on the boundary of one of these triangles and the corresponding interpolation formula was applied.

D. Generation of Synthetic Pollution Correlation Coefficients

The synthetic correlation coefficients for two-point pollutant concentrations were generated by program PSCOV which is reproduced on the pages following this discussion. The pages following the program contain a print-out of the input parameters that were used.

1. Main Program

The input parameters are read into the program in lines 15 through 87. The wind selector, JW, was used (line 20) to provide for wind information from either a frequency function (JW=1) or from frequency tables (JW=2) and provided a termination of the computation (JW=0). The subdivision of wind directions and speeds, stability classes, and inversion heights were read in on lines 29 through 35. The parameters of the wind frequency function are read on lines 45 through 53 (option JW=7, that actually used). The provision for the wind frequency tables is contained in lines 55 through 62. The ground coordinates of the locations at which the correlation coefficients are to be computed are read on line 67. Pollution sources, strength, effective stack height, and coordinates are read on line 82. A minimum pollutant concentration was read on line 86. The purpose of this was to provide a very small, but non-zero pollutant concentration at all points in all situations so as to avoid the possibility of an impossible logarithm of pollutant concentration which might occur in subsequent calculations.

The accumulation of data is initialized on line 90. The data accumulation is carried out in a sequence of nested DO-loops beginning on line 96 and ending on line 170. Subroutines FREQF (line 112) and SIGMAS (line 152) are called in this section and are discussed subsequently.

The statistical parameters are accumulated and reduced in the section from lines 174 through line 190. The section from line 190 to the end consists of output instructions. Subroutine MAP1 is used to printout the correlation coefficients, covariances, and other parameters in a "map" format so that contours of equal correlation coefficient (for example) could be rapidly drawn.

2. Subroutines

a) Subroutine FREQF

The wind frequency function (probability density) is a combination of circularly normal bivariate distributions

$$f(v,\theta) = k_1 f(v,\theta;\overline{w}_1,\overline{\theta}_1,\sigma_1) + \dots + k_n f(v,\theta;\overline{w}_n,\overline{\theta}_n,\sigma_n)$$

in which $k_1 + \dots + k_n = 1$, the k_i's being the fraction of the total wind population in the ith category and the parameters $\overline{w}_i$, $\overline{\theta}_i$, σ_i are the speed and direction of the mean wind vectors and the vector standard deviation of wind speed in the ith category. The probability density functions on the right differ only in the values assigned to the parameters. Thus

$$f(v,\theta;\bar{w},\bar{\theta},\ )=(\pi\sigma^2)^{-1}\exp\{-\sigma^{-2}[v^2+\bar{w}^2-2v\bar{w}\cos(\theta-\bar{\theta})]\}$$

The common multiplier for all the density functions, $vdvd\theta$, is omitted in the above. It has been found that the most complex wind distributions can be represented succinctly and with reasonable accuracy in the above form. This saves storage of extensive tables of observed frequency distributions.

b) Subroutine SIGMAS

This subroutine provides for the computation of the "sigmas" that appear in the pollution concentration formula. The data were taken from E.C. Eimutis and M.G. Konicek, Derivations of Continuous Functions for the Lateral and Vertical Atmospheric Dispersion Coefficients, Atmospheric Environment, Pergamon Press, 1972, Vol. 6, pp. 859-863.

c) Subroutine MAP

This subroutine prints out on the standard page printer the data input to it at the proper coordinates (as nearly as the printer will allow). Since we are primarily interested in correlation coefficients, the decimal point serves not only to fix the magnitude of the quantity concerned, but also as a "fix" for the point location.

```
      PROGRAM PSCOV(INPUT,OUTPUT,TAPE2)
C     COMPUTES MEANS,STD.DEV.,COV.,AND COR. OVER POINTS FROM
C     UP TO 30 POINT SOURCE INPUTS.
C
      COMMON                  F(18,10,6),X(100),Y(100),QS(30),XS(30),
     *        YS(30),HS(30),CBAR(100),CCOV(5050),CHITL(100),CSIG(100),
     *        CCOR(5050)
      COMMON/FFF/IW(6),WK(6,4),WBAR(6,4),TBAR(6,4),WSIG(6,4)
     *          ,DIR(18),VEL(10),ID,IV
      COMMON/STAB/AL(6),ISTAB(6)
      DIMENSION KNT(10),ARRAY(100),ITYPE(4)
      DIMENSION XI(100),YI(100)
      PI=3.1415926535
      RAD=0.0174533
C
C     INPUT QUANTITIES
C     JW=WIND DATA SELECTOR, JW=0 FOR STOP,JW=1 FOR FREQ. FUNCTION,
C     JW=2 FOR TABLES
C
   10 READ 1005,JW
 1005 FORMAT(I5)
      IF(JW.EQ.0) STOP
C
C     ID=NO.OF WIND DIRECTIONS, DIR(I)=DIRECTION TABLE, MAX=18
C     IV=NO.OF WIND SPEEDS, VEL(I)=SPEED TABLE, MAX=10
C     IST=NO. OF STABILITY TYPES, ISTAB(I)=STABILITY TABLE,
C     AL(I)=INVERSION HEIGHT, MAX = 6, 0. = NO INVERSION
C
      READ 1010,ID,(DIR(I),I=1,ID)
      READ 1010,IV,(VEL(I),I=1,IV)
 1010 FORMAT(I5/(8F8.2))
      READ 1015,IST
 1015 FORMAT(I5)
      READ 1116,(ISTAB(I),AL(I),I=1,IST)
 1116 FORMAT( 5(I5,F10.2))
      GO TO (15,25),JW
C
C     READS FREQ. FUNCTION DATA
C     IW(I), I=1,IST, NO. OF FREQ. FUNCTIONS FOR EACH STABILITY CASE
C     WK(I)=FRACTION DUE TO ITH FUNCTION
C     WBAR(I)=MEAN SPEED FOR ITH FUNCTION
C     TBAR(I)=MEAN DIRECTION
C     WSIG(I)=VECTOR STD.DEV.
C
   15 READ 1016,(IW(I),I=1,IST)
 1016 FORMAT(6I5)
      DO 20 I = 1,IST
      I1=IW(I)
      DO 20 J=1,I1
      READ 1020,WK(I,J),WBAR(I,J),TBAR(I,J),WSIG(I,J)
   20 CONTINUE
 1020 FORMAT(4F10.2)
      GO TO 35
C
C     READS FREQUENCY TABLES
```

```
C      F(I,J,K)=FREQUENCY TABLE ENTRIES
C
   25 DO 30 I=1,IST
      DO 30 J=1,ID
      READ 1025,(F(J,K,I),K=1,IV)
   30 CONTINUE
 1025 FORMAT(8F10.0)
C
C      READS GROUND COORDINATES
C      IG=NO. OF GROUND POINTS,X(I),Y(I)=COORDINATES
C
   35 READ 1030,IG,(XI(I),YI(I),I=1,IG,9)
 1030 FORMAT(I5/(2F10.0))
      II=0
      DO 36 I=1,IG,9
      DO 36 J=1,IG,9
      II=II+1
      X(II)=XI(I)
   36 Y(II)=YI(J)
C
C      READS SOURCE POINTS AND PARAMETERS
C      IS=NO. OF SOURCE POINTS
C      QS(I)=SOURCE STRENGTH
C      XS(I), YS(I)=SOURCE COORDINATES
C      HS(I)=EFFECTIVE SOURCE HEIGHT
C
      READ 1040,IS,(QS(I),XS(I),YS(I),HS(I),I=1,IS)
 1040 FORMAT(I5/(4F10.0))
C
C      CHIMIN=MINIMUM POLLUTANT CONCENTRATIONS
      READ 1035,CHIMIN
 1035 FORMAT(E10.4)
C
C      INITIALIZES STATISTICAL SUMS
      DO 50 I=1,5050
   50 CBAR(I)=0.
      SUMF=0.
C
C      LOOP ON STABILITY CLASSES
C
      DO 115 I1=1,IST
      ISTB=ISTAB(I1)
      ALL=AL(I1)
C
C      LOOP ON WIND DIRECTION
C
      DO 115 I2=1,ID
      AA=(270.-DIR(I2))*RAD
      CC=COS(AA)
      SS=SIN(AA)
C
C      LOOP ON WIND SPEED
C
      DO 115 I3=1,IV
      V=VEL(I3)
```

```
      GO TO (55,60) JW
   55 CALL FREQF(I1,I2,I3,FF)
      GO TO 65
   60 FF=F(I2,I3,I1)
   65 IF(FF.EQ.0.) GO TO 115
      SUMF=SUMF+FF
C
C     LOOP ON COORDINATES
C
      INDEX=0
      DO 110 I4=1,IG
      XX=X(I4)
      YY=Y(I4)
C
C     ROTATES TO WIND COORDINATES
C
      X1=XX*CC+YY*SS
      Y1=-XX*SS+YY*CC
C
C     LOOP ON SOURCES
C
      SUM=CHIMIN
      DO 100 I5=1,IS
      XXS=XS(I5)
      YYS=YS(I5)
C
C     ROTATES TO WIND COORDINATES
C
      XS1=XXS*CC+YYS*SS
      YS1=-XXS*SS+YYS*CC
C
C     RELATIVE WIND COORDINATES
C
      XR1=X1-XS1
      YR1=Y1-YS1
C
C     COMPUTES STRENGTH FROM EACH SOURCE
C
      IF(XR1.GT.0.) GO TO 95
      CHI=0.
      GO TO 100
   95 CALL SIGMAS(I1,ISTB,XR1,SIGY,SIGZ,H)
      CHI=QS(I5)/(PI*SIGY*SIGZ*V)
      CHI=CHI*EXP(-.05*((YR1*1000./SIGY)**2+(H*HS(I5)/SIGZ)**2))
  100 SUM=SUM+CHI
      CHITL(I4)=ALOG10(SUM)
C
C     CHITL(I4)=LOGE TOTAL CONCENTRATICN AT POINT I4 DUE TO
C     STAB, CLASS I1, WIND DIR I2, WIND SPEED I3
C     NOW COMPILE STATISTICS
C
      CBAR(I4)=CBAR(I4)+CHITL(I4)*FF
      DO 105 I6=1,I4
      INDEX=INDEX+1
  105 CCOV(INDEX)=CCOV(INDEX)+CHITL(I4)*CHITL(I6)*FF
```

```
C
C     END LOOP ON COORDINATES
  110 CONTINUE
C     END LOOP ON SPEED, DIR., AND STABILITY
  115 CONTINUE
C
C     PUTS STATISTICS IN STANDARD FORM AND UNITS
C
      IDX=0
      DO 120 I=1,IG
      CBAR(I)=CBAR(I)/SUMF
      DO 120 J=1,I
      IDX=IDX+1
  120 CCOV(DX)=CCOV(IDX)/SUMF-CBAR(I)*CBAR(J)
      IDX=0
      DO 130 I=1,IG
      DO 130 J=1,I
      IDX=IDX+1
      IF(J.EQ.1) CSIG(I)=SQRT(CCOV(IDX))
  130 CONTINUE
      IDX=0
      DO 140 I=1,IG
      DO 140 J=1,I
      IDX=IDX+1
  140 CCOR(IDX)=CCOV(IDX)/(CSIG(I)*CSIG(J))
C
C     PRINT OUT OF INPUT DATA
      PRINT 2000
 2000 FORMAT(1H1)
      PRINT 2005
 2005 FORMAT(* INPUT DATA FOR POINT SOURCE STATISTICS*)
      PRINT 2010,ID
 2010 FORMAT(*0WIND DIRECTIONS*,I5)
      PRINT 2015,(DIR(I),I=1,ID)
 2015 FORMAT(10F10.2)
      PRINT 2020,IV
 2020 FORMAT(*0WIND SPEEDS*,I5)
      PRINT 2025,(VEL(I),I=1,IV)
 2025 FORMAT(10F10.2)
      PRINT 2030,IST
 2030 FORMAT(*0STABILITY CLASSES*,I5)
      PRINT 2035,(ISTAB(I),I=1,IST)
 2035 FORMAT(6I5)
      GO TO (200,210) JW
  200 PRINT 2040
 2040 FORMAT(*0WINDS FROM CIRCULAR DISTRIBUTIONS*/5X*STAB.*5X*FRAC-*
     16X*MEAN*6X*MEAN*4X*VECTOR*/5X*CLASS*6X*TION*6X*DIR.*6X*VEL.*
     22X*STD.DEV.*)
      DO 205 I=1,IST
      I1=IW(I)
      DO 205 J=1,I1
      PRINT 2045,ISTAB(I),WK(I,J),TBAR(I,J),WBAR(I,J),WSIG(I,J)
  205 CONTINUE
 2045 FORMAT(I10,4F10.2)
      GO TO 220
```

```
  210 PRINT 2050
 2050 FORMAT(*0WINDS FROM FREQUENCY TABLES*)
      DO 215 I=1,IST
      PRINT 2055,ISTAB(I)
 2055 FORMAT(*0STABILITY CLASS*,I5)
      PRINT 2060,(VEL(J),J=1,IV)
 2060 FORMAT(8X,10F8.2)
      DO 215 K=1,ID
      PRINT 2065,DIR(K),(F(J,K,I),J=1,IV)
 2065 FORMAT(F8.2,10F8.2/(8X,10F8.2))
  215 CONTINUE
  220 PRINT 2070
 2070 FORMAT(*0GROUND COORDINATE PAIRS*/4(5X*POINTS*9X*X*9X*Y*) )
      DO 225 I=1,IG,4
      I1=I  $  I2=I+3
      IF(I2.GE.IG) I2=IG
      PRINT 2075,(J,X(J),X(J),J=I1,I2)
  225 CONTINUE
 2075 FORMAT(4(I10,2F10.0))
      PRINT 2080
 2080 FORMAT(*0SOURCE STRENGTH PARAMETERS*/4X*POINT*2X*STRENGTH*8X*XS*
     18X*YS*8X*HS*)
      DO 230 I=1,IS
      PRINT 2085,I,QS(I),XS(I),YS(I),HS(I)
  230 CONTINUE
 2085 FORMAT(I7, E13.4,3F10.0)
      PRINT 2090,CHIMIN
 2090 FORMAT(*0MINIMUM CONCENTRATION =*,1PE10.3)
      PRINT 2000
C     PRINTS OUT RESULTS
C
C
      PRINT 2100
 2100 FORMAT(* MEAN LOG10 CONCENTRATION BY GROUND POINTS*)
      DO 240 I=1,IG,8
      I1=I
      I2=I+7
      IF(I2.GE.IG) I2=IG
      PRINT 2105,(J,CBAR(J),J=I1,I2)
  240 CONTINUE
 2105 FORMAT(8(I3,1PE12.4))
      PRINT 2110
 2110 FORMAT(*0STD.DEV. OF LOG10 CONCENTRATION BY GROUND POINTS*)
      DO 245 I=1,IG,8
      I1=I
      I2=I+7
      IF(I2.GE.IG) I2=IG
      PRINT 2105,(J,CSIG(J),J=I1,I2)
  245 CONTINUE
      GO TO 270
      PRINT 2000
      PRINT 2115
 2115 FORMAT(*0COVARIANCE MATRIX*)
      IDX=0
      DO 250 I=1,IG
```

PROGRAM PSCOV

```
      DO 250 J=1,I,8
      J1=J
      J2=J+7
      IF(J2.GT.I) J2=I
      K=0
      DO 248 JJ=J1,J2
      IDX=IDX+1
      K=K+1
      KNT(K)=JJ
  248 ARRAY(K)=CCOV(IDX)
      PRINT 2119,(KNT(KK),KK=1,K)
 2119 FORMAT(5X,8(5X,I10))
  250 PRINT 2120,I,        ( ARRAY(KK),KK=1,K)
 2120 FORMAT(I5,8(3X,1PE12.4))
      PRINT 2000
      PRINT 2125
 2125 FORMAT(*0CORRELATION COEFFICIENT MATRIX*)
      IDX=0
      DO 265 I=1,IG
      DO 265 J=1,I,10
      J1=J
      J2=J+9
      IF(J2.GE.I) J2=I
      K=0
      DO 263 JJ=J1,J2
      IDX=IDX+1
      K=K+1
      KNT(K)=JJ
  263 ARRAY(K)=CCOR(IDX)
      PRINT 2149,(KNT(KK),KK=1,K)
 2149 FORMAT(5X,10(5X,I6))
  265 PRINT 2150,I,         (ARRAY(KK),KK=1,K)
 2150 FORMAT(I5,10(5X,F6.3))
C
C     PLOT STATISTICS
C
  270 ITYPE(1)=10HMEAN LOG10
      ITYPE(2)=10H CONCENTRA
      ITYPE(3)=10HTION
      ITYPE(4)=10H
      CALL MAP1(IG,X,Y,CBAR,1,ITYPE)
      ITYPE(1)=10HSTD.DEV. O
      ITYPE(2)=10HF LOG10 CO
      ITYPE(3)=10HNCENTRATE
      CALL MAP1(IG,X,Y,CSIG,2,ITYPE)
      GO TO 335
      ITYPE(1)=10HCOVARIANCE
      ITYPE(2)=10H MATRIX
      IDX=0
      DO 330 I=1,IG
      K=1
      DO 330 J=1,I
      IDX=IDX+1
      ARRAY(K)=CCOV(IDX)
  300 K=K+1
```

```
      IND=IDX
      INC=I
      I1=I+1
      DO 320 L=I1,IG
      ARRAY(K)=CCOV(IND+INC)
      IND=IND+INC
      INC=INC+1
  320 K=K+1
      ENCODE(20,3000,ITYPE(3)) I
 3000 FORMAT(*LOCATION*,I3,9X)
      CALL MAP1(IG,X,Y,ARRAY,2,ITYPE)
  330 CONTINUE
  335 ITYPE(1)=10HCORR.COEFF
      ITYPE(2)=10H. MATRIX
      IDX=0
      DO 360 I=1,IG
      K=1
      DO 340 J=1,I
      IDX=IDX+1
      ARRAY(K)=CCOR(IDX)
  340 K=K+1
      IND=IDX
      INC=I
      I1=I+1
      DO 350 L=L1,IG
      ARRAY(K)=CCOR(IND+INC)
      IND=IND+INC
      INC=INC+1
  350 K=K+1
      ENCODE(20,3000,ITYPE(3)) I
      CALL MAP1(IG,X,Y,ARRAY,2,ITYPE)
  360 CONTINUE
      GO TO 10
      END
```

SUBROUTINE FREQF

```
      SUBROUTINE FREQF(I1,I2,I3,FF)
      COMMON/FFF/IW(6),WK(6,4),WBAR(6,4),TBAR(6,4),WSIG(6,4)
     *          ,DIR(18),VEL(10),ID,IV
      DATA RAD/.0174532925/
      V1=VEL(I3)
      D1=DIR(I2)
      IF(I2.EQ.1) 2,4
    2 DD=(DIR(2)+360.-DIR(ID))/360.
      GO TO 10
    4 IF(I2.EQ.ID) 6,8
    6 DD=(DIR(1)+360.-DIR(ID-1))/360.
      GO TO 10
    8 DD=(DIR(I2+1)-DIR(I2-1))/360.
   10 IF(I3.EQ.1) 12,14
   12 DV=(VEL(1)+VEL(2))/2.
      GO TO 20
   14 IF(I3.EQ.IV)16,18
   16 DV=1.5*(VEL(IV)-VEL(IV-1))
      GO TO 20
   18 DV=(VEL(I3+1)-VEL(I3-1))/2.
   20 J2=IW(I1)
      SUM=0.
      DO 22 J=1,J2
      SIGSQ= WSIG(I1,J)**2
      W1=WBAR(I1,J)
      F=(V1**2+W1**2-2.*V1*W1*COS(RAD*DIR(I2)-RAD*TBAR(I1,J)))/SIGSQ
      F=WK(I1,J)*DD*DV*V1*EXP(-F)/SIGSQ
   22 SUM=SUM+F
      FF=SUM
      RETURN
      END
```

SUBROUTINE SIGMAS

```
      SUBROUTINE SIGMAS(I1,ISTB,X,SIGY,SIGZ,H)
      DIMENSION A(6),A1(18),B1(18),C1(18)
      COMMON/STAB/AL(6),ISTAB(6)
      DATA (A(I),I=1,6)/.3658,0.2751,0.2089,0.1471,0.1046,0.0722/
      DATA (A1(I),I=1,18) /0.00024,0.055,0.113,1.26,6.73,18.05,
     *                     0.0015,0.028,0.113,0.222,0.211,0.086,
     *                     0.192,0.156,0.116,0.079,0.063,0.053/
      DATA (B1(I),I=1,18) /2.094,1.098,0.911,0.516,0.385,0.180,
     *                     1.941,1.149,0.911,0.725,0.678,0.740,
     *                     0.936,0.922,0.905,0.881,0.871,0.814/
      DATA (C1(I),I=1,18) /+9.6,2.0,0.0,-13.,-34.,-48.6,
     *                     9.27,3.3,0.0,-1.7,-1.3,-0.35,
     *                     6*10.0/
      X=1000.*X
      SIGY=A(ISTB)*(X**0.9031)
      IF(X.GT.1000.)2,4
    2 J=ISTB
      GO TO 18
    4 IF(X.LT.100.) 6,8
    6 J=ISTB+12
      GO TO 10
    8 J=ISTB+6
   10 IF(AL(I1).GT.0.) 12,16
   12 XL=((0.47*AL(I1)-C1(J))/A1(J))**(1.0/B1(J))*2.
      IF(X.GT.XL)14,16
   14 SIGZ=0.04
      H=0.
      GO TO 18
   16 SIGZ=A1(J)*(X**B1(J))+C1(J)
      H=1.
   18 RETURN
      END
```

SUBROUTINE MAP

```
      SUBROUTINE MAP(NP,X,Y,MATRIX,ISW,KIND)
      COMMON/M/ARRAY(13,58),LINE(100),NSPACE(100)
      DIMENSION X(100),Y(100),MATRIX(100),KIND(4)
      REAL MATRIX
      GO TO (10,100) ISW
   10 DO 15 I=1,58
      DO 15 J=1,13
   15 ARRAY(J,I)=10H
C
C     FIND PRINTER LOCATION OF X AND Y COORDINATES
C
      XMIN=X(1)
      YMIN=Y(1)
      YMAX=Y(1)
      DO 20 N=2,NP
      XMIN=MIN1(X(N),XMIN)
      YMIN=MIN1(Y(N),YMIN)
      YMAX=MAX1(Y(N),YMAX)
   20 CONTINUE
      YF=48./(YMAX-YMIN)
      XF=80./(YMAX-YMIN)
      DO 30 N=1,NP
      NSPACE(N)=(X(N)-XMIN)*XF+1.
      LINE(N)=(YMAX-Y(N))*YF+2.
   30 CONTINUE
C
C     PUT MATRIX VALUES INTO PROPER PRINTER LOCATIONS
C
  100 DO 200 I=1,NP
      NWORD=NSPACE(I)/10
      NCHAR=NSPACE(I)-NWORD*10
      N=NWORD+2
      L=LINE(I)
      IF(NCHAR.GT.0) GO TO 110
      ENCODE(10,1010,ARRAY(N,L)) MATRIX(I)
 1010 FORMAT(1H*F6.3,3X)
      GO TO 200
  110 GO TO (121,122,123,124,125,126,127,128,129) NCHAR
  121 ENCODE (10,1011,ARRAY(N,L)) MATRIX(I)
 1011 FORMAT(X,1H*,F6.3,2X)
      GO TO 200
  122 ENCODE (10,1012,ARRAY(N,L)) MATRIX(I)
 1012 FORMAT(2X,1H*,F6.3,X)
      GO TO 200
  123 ENCODE(10,1013,ARRAY(N,L)) MATRIX(I)
 1013 FORMAT(3X,1H*,F6.3)
      GO TO 200
  124 ENCODE(20,1014,ARRAY(N,L)) MATRIX(I)
 1014 FORMAT(4X,1H*,F6.3,9X)
      GO TO 200
  125 ENCODE(20,1015,ARRAY(N,L)) MATRIX(I)
 1015 FORMAT(5X,1H*,F6.3,8X)
      GO TO 200
  126 ENCODE(20,1016,ARRAY(N,L)) MATRIX(I)
 1016 FORMAT(6X,1H*,F6.3,7X)
```

SUBROUTINE MAP

```
          GO TO 200
      127 ENCODE(20,1017,ARRAY(N,L)) MATRIX(I)
     1017 FORMAT(7X,1H*F6.3,6X)
          GO TO 200
      128 ENCODE(20,1018,ARRAY(N,L)) MATRIX(I)
     1018 FORMAT(8X,1H*F6.3,5X)
          GO TO 200
      129 ENCODE(20,1019,ARRAY(N,L)) MATRIX(I)
     1019 FORMAT(9X,1H*,F6.3,4X)
      200 CONTINUE
          WRITE(2,2002) KIND
     2002 FORMAT(*1*,4A10)
          DO 210 L=1,58
          WRITE(2,2004)(ARRAY(N,L),N=1,13)
      210 CONTINUE
     2004 FORMAT(X,13A10)
          RETURN
          ENTRY MAP1
          II=0
          IS=9
          DO 330 N=1,NP,9
          I=IS
          DO 320 NN=1,9
          II=II+1
          X(II)=MATRIX(I)
      320 I=I+9
          IS=IS-1
      330 CONTINUE
          PRINT 2002,KIND
          PRINT 2006,(X(I),I=1,NP)
     2006 FORMAT(9F10.3/////)
          RETURN
          END
```

(Input data on point source statistics is listed in Table III-1, p. 33)

CHAPTER IV

THE ANALYSIS OF THE COVARIANCE MATRICES

In the use of the covariance function or correlation coefficient function it is important to evaluate the part that may be due to small scale effects and to random errors. The problem is analogous to the problem in communication theory in which a signal is observed in a noisy background. In order to determine the signal, it is also required that a considerable amount of information be available concerning the nature of the noise.

The situation in general was described in qualitative terms in Chapter I where the effect of a jump discontinuity at zero distance or a part with small "range of influence" on the statistical interpolation were discussed. Up to this point the method of finding the magnitude of this discontinuity at zero distance or the magnitude of the effects with small "range of influence" have not been discussed. They are the specific subject of this chapter.

A. Evaluation of the Effect of Errors of Measurement or of Small Scale Phenomena

To evaluate the effect of errors of measurement or of small scale phenomena, we carry further the procedures that led to the mean square estimate of Chapter II, equation (8). The expression for the estimate of pollution concentration, $\hat{y}$, at P was given in (9), Chapter II, as

$$\hat{y} = \sum_i x_i \left(\sum_j a^{ij} g_j \right) \qquad (1)$$

where the x_i's are the observed concentration measures on a particular occasion, $g_j = \overline{(x_j y)}$ is the covariance of the concentration measures at P_j and P, a^{ij} is an element (row i, column j) of the inverse of the covariance matrix $\{a_{ij}\}$,

$a_{ij}=\overline{(x_i x_j)}$, where a_{ij} is the covariance of concentration measures at P_i and P_j. It was pointed out following equation (9) of Chapter II that if we let $P \to P_k$, an observation point, and assume that $g_j \to a_{jk}$ at the same time, then $\hat{y} \to x_k$, i.e., the estimate at P approaches the observed value at P_k when P approaches P_k.

We consider now the situation illustrated in Figures I-3a, I-3b and Figures I-5a, I-5b. To express the ideas shown there qualitatively we need an explicit formulation for the covariances $g_j=g_j(P_j,P)$. Thus, let

$$g_j = g_j^{(1)} + g_j^{(2)}$$

where $g_j^{(1)}$ is the part of the covariance g_j which describes the overall variation of g_j as a function of the location of P with respect to P_j while $g_j^{(2)}$ is the part that represents the amount of discontinuity in g_j at P_j on the one hand or the part of g_j that has a limited "range of influence". The expression "limited" also needs definition (or at least needs to be made explicit). For our purposes, "limited range of influence" will be taken to mean that at distances from P_j to P that are of the order of magnitude of the spacing between prospective observation sites the magnitude of $g_j^{(2)}$ is negligible (i.e., the presence of $g_j^{(2)}$ cannot be distinguished from sampling variations). With this specification of "limited range of influence", the situation is treated as though $g_j^{(2)}$ were simply the magnitude of the jump discontinuity if $g_j(P_j,P)$ for $P \to P_j$. Then we may write

$$g_j(P_j,P_j) = g_j^{(1)}(P_j,P_j) + g_j^{(2)}(P_j,P_j)$$

$$g_j(P_j,P) = g_j^{(1)}(P_j,P), \quad P \neq P_j$$

or

$$g_j^{(2)}(P_j,P) = 0 \text{ if } P \neq P_j$$

$$g_j^{(2)}(P_j,P) = g_j^{(2)}(P_j,P_j) \text{ if } P = P_j$$

Equation (1) then becomes

$$\hat{y} = \sum_i x_i \left(\sum_j a^{ij} g_j^{(1)}\right), \quad P \neq P_j$$

$$\hat{y} = x_k, \quad P = P_k$$

where P_k may be any one of the points P_j.

Consider now the element of the covariance matrix $\{a_{ij}\}$, $a_{ij} = \overline{(x_i x_j)}$. Note that these covariances are essentially the same as those of $g_j = \overline{(x_j y)}$ except for the fact that the wandering point P involved in $g_j = g_j(P_j,P)$ is now restricted to one of the observation sites. As long as the points P_i and P_j at which the covariance $a_{ij} = \overline{(x_i x_j)}$ is computed are distinct, the values concerned are just those of $g_j^{(1)}(P_j,P_i)$. The part of the covariance that had "limited range of influence" is not involved. But throughout the principal diagonal of $\{a_{ij}\}$ one has i=j. For these elements of the matrix $\{a_{ij}\}$ both parts are involved. These elements of the matrix will be written as

$$a_{ii} = a_{ii}^{(1)} + \overline{e_i^2}$$

where $a_{ii}^{(1)}$ is what we have called $g_i^{(1)}(P_i,P_i)$ above and

$\overline{e_i^2}$ is what we have called $g_i^{(2)}(P_i,P_i)$. (Note: the quantity $\overline{e_i^2}$ is not to be confused with the mean square error of estimate, $\overline{e^2}$, without subscript, used in equation (8), Chapter II.) The terms $\overline{e_i^2}$ will be referred to as the "residual variances". They are the quantities that must be determined in order to separate the part $g_j^{(2)}$ from g_j.

The method used to determine the residual variances is an important part of Factor Analysis. It is unfortunate that in the physical sciences the importance of these residual variances has been, until recently, neglected to a large extent. At least a part of this is due to the fact that physical scientists also tend to be instrument designers and have felt that to recognize "errors of measurement" was to cast doubt on the quality of their instruments. Factor Analysis is traceable to the psychologist Carl Spearman (1904) and the recognition of the importance of the residual variances originates with him also. As used by psychologists, the covariances (or correlation coefficients) that make up the matrix elements a_{ij} are those between "test i" and "test j" when given to a group of subjects. The psychologists, like the physical scientists, were hesitant to admit that these tests involved "errors of measurement". They did, however, recognize that each test had something about it that was unique (belonged to it alone). Consequently what we here call the residual variances were called by the psychologists the "uniquenesses".

Factor Analysis, as such, is devoted to things quite different from the determination of the residual variances. These are rather incidental, but have an important bearing on the prime objective in that subject. We confine our attention to finding the residual variances alone and ignore all other aspects of Factor Analysis. The best treatment that we have found is that of Lawley and Maxwell (1963) and (1971) who treat the subject from the point of view of a statistician and include discussions of the applicable tests of significance. The ordinary texts on Factor Analysis tend to emphasize the computational and interpretive aspects of the subject and neglect the significance tests. For example, Horst (1965) is a 730 page compendium of computing programs and the associated mathematical manipulations which does not mention a single test for significance. On the other hand, adequate tests of significance only go back as far as Bartlett (1951) and Lawley (1956). The methods used here depend chiefly on Jöreskog (1962) and may also be found in Lawley and Maxwell (1971) [but not in (1963)]. As far as we have been able to determine, the treatment of the method of finding the residual variances using the formulation of the problem as an integral equation rather than as a matrix equation has not appeared elsewhere except for a brief note by the writer (Buell, 1972) which covers only a small part of the work reported here.

B. Determination of the Residual Variances

The discussion of Section A preceding leads us now to consider the matrix $A = \{a_{ij} + \overline{e_i^2}\delta_{ij}\}$, where $\delta_{ij}=0$ if $i\neq j$ and $=1$ if $i=j$ so that we may write

$$A = \begin{pmatrix} a_{11}+\overline{e_1^2}, & a_{12} & , & ---, & a_{1n} \\ A_{21} & , \; a_{22}+\overline{e_2^2} & , & ---, & a_{2n} \\ & --- & --- & & \\ a_{n1} & , \; a_{n2} & , & ---, & a_{nn}+\overline{e_n^2} \end{pmatrix}$$

and in which $a_{ij}=a_{ji}$, i.e., the matrix A is symmetric. The matrix A is also positive definite. To keep the notation from becoming excessivly complex, the diagonal terms shown as a_{ii} here were denoted by $a_{ii}^{(1)}$ in the previous section. We note now that one may write the matrix A as a sum of two matrices, A=C+D, where C is the matrix $\{a_{ij}\}$ and D is the diagonal matrix $\{\overline{e_j^2}\delta_{ij}\}$. The problem at hand consists of finding the elements of D and the diagonal elements of C (i.e., the values of a_{ii}) so that their sum, $a_{ii} + \overline{e_i^2}$ will have the known value that appears on the diagonal of the given (known) matrix A and such that all of the diagonal elements of D are positive (or perhaps zero) and such that the resulting matrix C will be positive definite.

1. Representation of a Matrix in Terms of Proper Functions/ Proper Values

In this section some background material is introduced which forms the basis on which the following sections depend. We consider a symmetric positive definite matrix $B=\{b_{ij}\}$. It is well known that such a matrix may be written in the form

$$B=\Phi\Lambda\Phi', \qquad \Phi'= \text{transpose of } \Phi \qquad (2)$$

in which the matrix Λ is a diagonal matrix such that the elements on the principal diagonal, λ_i, are all positive (or at least non-negative) real numbers. It is also further specified that these be written in decreasing order of magnitude:

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq --- \geq \lambda_n \geq 0$$

These are the proper values belonging to the matrix B. (Also called eigenvalues, latent roots, characteristic roots, etc.) The matrix B is assumed to have n rows and columns. There are then n proper values, not necessarily all distinct. The proper values are the solutions of the determinantal equation

$$|B-\lambda I| = 0$$

where $|\cdot|$ stands for the determinant and I is the nXn unit matrix. This equation is a polynomial of degree n in λ.

With each of the proper values, there is associated a proper function (or proper vector) which appears in the column of Φ that corresponds to the position of the associated proper value in Λ. Thus to λ_i, there corresponds the column vector col$\{\phi_{1i},\phi_{2i}, ---,\phi_{ni}\}$ which stands in the i'th column of Φ. (These are also called eigenvectors, latent vectors, characteristic vectors, etc.) These are the solutions of the homogeneous equations

$$\sum_k b_{jk}\phi_{ki} = \lambda_i\phi_{ki} \qquad j=1,---,n \qquad (3)$$

(It is also said that the matrices Λ, Φ, are the solutions of the matrix equation $B\Phi=\Phi\Lambda$.)

The proper vectors or proper functions have the property of being an orthonormal set of vectors. That is to say, they are normalized,

$$\sum_k \phi_{ki}{}^2=1, \qquad i=1,---,n \qquad (4)$$

and they are orthogonal to each other

$$\sum_k \phi_{ki}\phi_{kj}=0, \qquad i\neq j. \qquad (5)$$

The original statement that the matrix B may be expressed in terms of its proper values and proper functions is written in explicit summation notation as

$$b_{ij} = \sum_k \lambda_k\phi_{ik}\phi_{jk}. \qquad (6)$$

The decomposition of a matrix into its proper values/ proper vectors is an important aspect of matrix analysis, of which we have presented only the bare essentials for a particular case. See any standard text such as Bellman (1960), Gantmacher (1960a), MacDuffee (1949), Perlis (1952), Turnbull (1960), etc. for a general treatment. Not only is it important for the subject at hand, but has many applications, particularly to oscillatory systems from simple strings and mechanical systems (Gantmacher, 1960b) to nuclear spectra (Mehta, 1967).

In the above, the terms proper vectors and proper functions have been used interchangeably. This usage is intentional and is to emphasize the fact that what is termed a "vector" with n discrete "components" in this discussion will turn out to be a function on a continuum and it just happens that our computing procedure is limited to the direct computation of values of this function at only n points.

2. Factor Analysis Methods

Modern factor analysis techniques provide methods for finding the individual values of $\overline{e_i^2}$, i=1, ---, n. Most of the methods for finding these quantities (i.e., those that are statistically sound) require the solution of the matrix equation

$$(A-D)\Phi = \Phi\Lambda \qquad (7)$$

where D is the diagonal matrix with elements $\overline{e_i^2}$ on the principal diagonal and 0's elsewhere. All of D,Λ,Φ are to be found while only A is given. Nearly all of the methods concerned are iterative, and have the characteristic feature that (with only a few exceptions) they converge very slowly (and in some cases, some methods fail to converge at all) (Lawley and Maxwell, 1963).

The method used here is straight-forward and has given results that appear to be satisfactory. It is that developed by K.G. Jöreskog (1962). The technique is as follows. It is assumed that the covariance matrix, A, may be written in the form

$$A = \Phi\Lambda\Phi' + D \tag{8}$$

where Φ' is the transpose of Φ which in turn is the matrix of proper functions of A-D, Λ is the diagonal matrix of proper values, and $D = \{\overline{e_i^2}\sigma_{ij}\}$ is the diagonal matrix of residual variances. The proper functions (or vectors) corresponding to the proper values are ordered in the same way as the associated proper values [$\mathrm{col}(\phi_{1i}, \phi_{2i}, ---, \phi_{ni})$ corresponding to λ_i]. The values of λ_i, i=1, ---, n, are the roots of the determinant equation

$$|A-D-\lambda I| = 0$$

where I is the n x n unit matrix. Now consider the matrix A-D instead of the matrix A and assume that all of the diagonal terms of D are identical, $\overline{e_i^2} = \overline{e^2}$, i=1,---,n. It is rather easily shown that (a) the matrix A has proper values which are those of A-D but increased by $\overline{e^2}$ and (b) the proper functions of A are exactly those of A-D. Thus, if μ_i, i=1,---,n, are the proper values of A, and λ_i, i=1,---,n, are those of A-D, then

$$\mu_i = \lambda_i + \overline{e^2} .$$

(Anderson, 1958, p. 287, problem 7)

If, now, we were to find the proper values of A as $\mu_1, \ldots, \mu_n$ and if we had some way of determining that the last n-k of these proper values were not significantly different from each other, it would then seem reasonable that we could lump these last n-k values into one batch with average value $(\mu_{k+1}, \ldots, +\mu_n)/(n-k)$, take this value as the estimate of $\overline{e^2}$, and then consider the values of $\mu_1-\overline{e^2}, \ldots, \mu_k-\overline{e^2}$ as proper values of the matrix A-D, which will then be of rank $k(<n)$ and will have only k proper functions, namely $\mathrm{col}(\phi_{1i}, \ldots \phi_{ni})$, $i=1,\ldots,k$, and the last n-k proper values, λ_i, would be zeros. It just so happens that just such a test as required above is available. Its specification is deferred until later since it is used in a somewhat more general sense than is required by the heuristic argument above.

Return now to the technique of Jöreskog (1962) in which the elements of $D=\{e_i^2\delta_{ij}\}$ may differ from each other. It is assumed that the matrix D is given by the expression $D = \theta[\mathrm{diag}(A^{-1})]^{-1}$, where θ is a scalar constant to be determined, D is proportional to the diagonal matrix which is the inverse of the diagonal of the inverse of the covariance matrix that we started with. We reduce this notation to the form $D=\theta\Delta^{-1}$, where $\Delta=\mathrm{diag}(A^{-1})$ and (8) then becomes

$$A = (\Phi\Lambda\Phi') + \theta\Delta^{-1}. \qquad (9)$$

Since A is a covariance matrix, the elements of Δ will all be positive, and if we take the positive square roots of the elements of Δ and denote this diagonal matrix by $\Delta^{1/2}$, it then follows that (9) may be written as

$$\Delta^{1/2}A\Delta^{1/2} = (\Delta^{1/2}\Phi\Delta^{1/2})(\Delta^{1/2}\Phi\Delta^{1/2})' + \theta I$$

or

$$\Delta^{1/2}A\Delta^{1/2} = CC' + \theta I$$

where

$$C = \Delta^{1/2}\Phi\Delta^{1/2}$$

The matrix $\Delta^{1/2}A\Delta^{1/2}$ is then to be expressed in the form described in the heuristic argument preceding where the single scalar coefficient, θ, corresponds to the common residual variance, $\overline{e^2}$, used there. To find the value of θ, one finds the proper functions and values of the matrix $\Delta^{1/2}A\Delta^{1/2}$, or

$$(\Delta^{1/2}A\Delta^{1/2})F^* = F^*\Lambda^*$$

where Λ^* is the required diagonal matrix of proper values (properly ordered) and F^* is the matrix with proper functions appearing in the columns in the same order as the proper values.

The test for the last n-k proper values being different from each other is to the effect that the quantity Q is approximately distributed as χ^2, where

$$Q = N'\{-\log_e(\lambda^*_{k+1}\cdot ---\cdot\lambda^*_n)+(n-k)\log[(\lambda^*_{k+1}+---+\lambda^*_n)/(n-k)]$$

and

$$N' = N-k+[2(n-k)+1+2/(n-k)]$$

and the number of degrees of freedom for χ^2 is

$$\text{d.f.} = (n-k+2)/(n-k-1)/2.$$

(N+1 = number of observations on which A is based.) The criterion operates in an inverse sense. If the value of χ^2 for a given value of k (the number of proper values accepted as being different from each other) is exceeded then there is at least one more significant proper value that should be included. If it is found that at a given significance level, the number of significantly different proper values, k, is adequate for the representation of the matrix $\Delta^{1/2}A\Delta^{1/2}$, then the value of θ is taken as

$$\theta = (\lambda_{k+1}^{*} + --- + \lambda_{n}^{*})/(n-k),$$

the average of the last (n-k) proper values.

We have now succeeded in representing the matrix $\Delta^{1/2}A\Delta^{1/2}$ in the form

$$\Delta^{1/2}A\Delta^{1/2} \cong F_k^{*}\Lambda_k^{*}(F_k^{*})' + \theta I$$

where Λ_k^{*}, F_k^{*} consist of only the first k proper values and proper functions. The matrix A may then be written as

$$A \cong (\Delta^{-1/2}F_k^{*})\Lambda_k^{*}(\Delta^{-1/2}F_k^{*})' + \theta\Delta^{-1}, \quad D=\theta\Delta^{-1}=\{\overline{e_i^{\,2}}\delta_{ij}\} \qquad (10)$$

Since the matrix $\Delta^{-1/2}F_k^{*}$ has columns not orthogonal to each other, these are no longer proper functions of any matrix. To obtain proper functions and values, it is necessary to recompute values of the k proper values/functions from scratch, only using as the matrix concerned the matrix $(\Delta^{-1/2}F_k^{*})\Lambda_k^{*}(\Delta^{-1/2}F_k^{*})'$. There are, of course, only k non-zero proper values and proper functions, Λ_k, Φ_k, which are usually quite close to those already obtained for Λ_k^{*}, F_k^{*} (except for a scale factor for the proper values).

To recapitulate, it was pointed out that the use of equation (9) of Chapter II as an interpolation formula for estimating the measure of pollutant concentration involves two important items that are frequently overlooked. The first is that the variance of the small scale effects, the effects of limited range of influence, must be specified quantitatively. The Jöreskog (1963) representation of the covariance matrix does this. The required values are exactly those specified by (10). The second item is that a method of interpolation was required that would preserve the character of the matrix A as a covariance matrix. This is provided when the first matrix expression on the right of (10) is recomputed in terms of its proper functions, Φ_k, and corresponding proper values, Λ_k so that

$$A \cong \Phi_k \Lambda_k \Phi_k' + D . \tag{11}$$

In (11) note that the subscript k is <u>not</u> a summation index. It is used to indicate that only the first k proper values and functions are used. The matrices concerned have dimensions as follows: Φ_k is (nXk), Λ_k is (kXk), Φ_k' is (kXn), so that $\Phi_k \Lambda_k \Phi_k'$ is (nXn) as are A and D.

It appears that the solution to the problems concerned is at hand, but this is not necessarily the case. The Factor Analysis method for finding the values (e^2) is applicable to quite general covariance matrices. We are concerned with covariance matrices of a stochastic process on a continuum. The Factor Analysis method is strongly dependent on the value used for the number of statistically significant proper values, k. The problem will next be considered using a continuum formulation. The effect on the evaluation of k is found to be important.

In Sub-section 3, Integral Equation Methods, immediately following, the problems involved in the solution of the integral equation corresponding to (7) are considered. A method for testing whether the solutions of this integral equation obtained by two (or more) only slightly different procedures are consistent with each other is developed. The section is strictly theoretical. The practical application of the techniques of Sub-section 3 is made in Sub-section 4 to actual SO_2 measurements.

3. Integral Equation Methods

The factor analysis method for determining the values of the residual variances fails to take into account the fact that there are strong geometric relations connecting the pollution concentration covariances at the various points of the observing network. To bring this into the picture, the problem is re-stated in terms of a continuum of values rather than in a totally discrete form that ignores this situation. To illustrate, the values x_i, x_j that enter into the covariances $\overline{(x_i x_j)}$ can be (in the factor analysis case frequently are) test scores from the i'th and the j'th tests given to a batch of subjects. The i'th test might be on arithmetic ability and the j'th test on manual dexterity. If a third test is considered, say x_k for the score on the k'th test, which is on social adaptiveness, it is silly to place this in a strictly geometrical relation with respect to the other two. On the other hand, when x_i, x_j, x_k are pollution concentration at points P_i, P_j, P_k, we know all about the strictly geometrical relations between these points, i.e., we can say that P_k is such and such a distance from P_i and also from P_j and that P_i and P_j are so far from each other and that the line joining them has a certain direction. The natural generalization of the proper value/function analysis of the covariance matrix is the integral equation formulation for proper values and functions in a continuum. Thus

$$\lambda\phi(x) = \int_R K(x,x')\phi(x')\, dx' \qquad (12)$$

where λ is a proper value, $\phi(x)$ is the corresponding proper function, and $K(x,x')$ is the kernel, the exact analogue of the covariance matrix. The kernel, $K(x,x')$ is precisely the covariance of concentrations at the points x and x'. Note that x in the above may be a multidimensional variable in which case dx' is multidimensional and R is a region of the same dimensions. We are concerned here with the two-dimensional field of pollution concentrations. The kernel function, $K(x,x')$, is now a positive definite symmetrical $[K(x,x') = K(x',x)]$ function of the location coordinates. The fact that we are now working on a continuum is also reflected in the fact that (12) has a denumerable infinity of solutions, λ_n, $\phi_n(x)$, of proper values and functions. The proper values are considered as ordered

$$\lambda_1 \geq \lambda_2 \geq --- \geq \lambda_n \geq --- \geq 0$$

and the proper functions $\phi_n(x)$ are taken in the same order.

The standard method of solving (12) is to reduce the problem to a matrix problem where the values x,x' are specified on a network of points and the integral is replaced by a quadrature formula. This approximation may be written as

$$\lambda\phi(x_i) \cong \sum_j K(x_i,x_j)\phi(x_j)A_j \tag{13}$$

where the factors A_j correspond to an element of area about the point x_j such that ΣA_j = R = area of the region of integration (for a two dimensional network of points). If the kernel function were given by an explicit formula, then one could evaluate it at any number of point pairs x_i and x_j, i,j=1, ..., n, and the number of points could be taken as very large. The larger the number of points selected, the more accurate would be the estimate of the solution for a proper value/function of some specific order, n. On the other hand, one is faced with the fact that one can only obtain a finite number of solutions when it is known in advance that there are a denumerable infinity of such. Thus, for a fixed number of points, x_i, i=1, ..., n, n fixed, there must always be some point, say n*, beyond which the solutions depart more and more from the theoretical or exact solution. When one applies the technique to a kernel function that is known only at point pairs from a fixed number of points and for which the value of the kernel function is known only to within a certain sampling error, other considerations limit the accuracy of the solutions. One of these is the accuracy of the values of the kernel function itself and another is the choice of the quadrature factors, A_j, and still a third is the fact that K(x,x') may have a "jump discontinuity" along x=x' (or at least an effective "jump discontinuity" as far as the observation station spacing is concerned). These several points will be considered separately in the following paragraphs. None of these points is trivial. Considered from the integral equation point of view, the proper values/functions obtained from the matrix equations by Factor Analysis methods are simply approximations to a correct solution in which several of the factors concerned have been overlooked due to the over-simplification of the problem.

a) The Jump Discontinuity

The fact that there may be a jump discontinuity along x=x' may be handled by the following integration technique. Subtract from the equation (12) the identity

$$\phi(x) \int_R K(x,x')dx' = \int_R K(x,x')\phi(x)dx'$$

to obtain

$$\phi(x)\,[\lambda-\int_R K(x,x')\,dx'] = \int_R K(x,x')\,[\phi(x')-\phi(x)]\,dx' \qquad (14)$$

The discontinuity of the kernel function in the integrand on the right of (14) is no longer effective since the factor $[\phi(x')-\phi(x)]$ is zero on $x=x'$. It is still present in the integral term on the left side. In this integral we use quadrature factors that are dependent on x in addition to x' in such a way that $B_j^{(x)}=0$ if $x=x_j$. Quadrature formulas using such factors are said to be of "open type". See Abramowitz and Stegun (1964). If these quadrature factors are indicated by $B_j^{(x)}$ while those on the right are A_j, then the matrix equivalent of (14) may be written as

$$\phi(x_i)\,[\lambda-\sum_j K(x_i x_j)B_j^{(x_i)}] = \sum_j K(x_i,x_j)\,[\phi(x_j)-\phi(x_i)]\,A_j \qquad (15)$$

Rearranging the terms of (15), one finds that

$$\lambda\phi(x_i)=\sum_{j\neq i} K(x_i,x_j)\,\phi(x_j)A_j+\sum_{j\neq i} K(x_i,x_j)\,\phi(x_i)\,[B_j^{(x_i)}-A_j] \qquad (16)$$

where neither summation contains the term $i=j$. This term is missing from the first summation because regardless of A_i, it would be canceled by the corresponding term of the second summation; it is missing from the second to also account for the fact that $B_i^{(x_i)} = 0$. The second term on the right is the equivalent of substituting for $K(x_i,x_j)A_i$ the expression

$$K(x_i,x_i)A_i=\sum_{i\neq j} K(x_i,x_j)\,[B_j^{(x_i)}-A_j]. \qquad (16a)$$

This is equivalent to using an interpolation formula to obtain $K(x_i,x_i)$ from values of $K(x_i,x_j)$ that make no use of $x_i = x_j$ explicitly. One must have $\sum_{j\neq 1} B_j^{(x_i)}=R$ and

$\sum_{j\neq i} A_j = R - A_i$ so that $\sum_{j\neq i} [B_j^{(x_i)} - A_j] = A_i$, i.e., the sum of the weights given to each of the terms $K(x_i,x_j)$ of (16a) is just exactly the weight ascribed to $K(x_i,x_i)$ itself.

In order to obtain an explicit solution using the symmetric positive definite matrix algorithms (which are much faster than those for an unsymmetric matrix), this form may be made symmetrical using the device described in the following section.

When the proper functions and values, $\phi_k^*(x_i), \lambda_k^*$, have been found from (16), then a new kernel $K^*(x_i,x_j)$ can be constructed, thus

$$K^*(x_i,x_j) = \sum_k \lambda_k^* \phi_k^*(x_i)\phi_k^*(x_j)$$

and the amount of the jump discontinuity along $x_i = x_j$ may be determined from

$$J_i = J(x_i) = K(x_i,x_i) - K^*(x_i,x_i).$$

b) <u>The Quadrature Factors</u>

The unsymmetrical formulation for the matrix approximation to the integral equation using a quadrature formula, (13), may be made completely symmetric by multiplying both sides by $\sqrt{A_i}$. Thus (13) becomes

$$\lambda[\sqrt{A_i}\phi(x_i)] = \{\sqrt{A_i}K(x_i,x_j)\sqrt{A_j}\}[\sqrt{A_j}\phi(x_j)] \tag{17}$$

If we let $\sqrt{A_i}\phi(x_i) = \theta(x_i)$, then the values of $\theta(x_i)$, λ are proper functions/values corresponding to the weighted covariance matrix $\{\sqrt{A_i}K(x_i,x_j)\sqrt{A_j}\}$. Note that if we have proper functions $\theta_k(x_i)$ and $\theta_l(x_i)$ corresponding to proper values λ_k,λ_l, these functions are orthonormal, i.e.,

$$\sum_i \theta_k(x_i)\theta_l(x_i) = \delta_{kl} \qquad \delta_{kl} = \begin{matrix} 1 \text{ if } k=l \\ 0 \text{ if } k\neq l \end{matrix}$$

If now we substitute the expressions for $\theta_k(x_i)$, $\theta_l(x_i)$ in terms of $\phi_k(x_i)$, $\phi_l(x_i)$, one obtains

$$\sum_i \phi_k(x_i)\phi_l(x_i)A_i = \delta_{kl}$$

which corresponds to the orthonormality condition in integral form which is satisfied by the proper functions belonging to the integral equation (12),

$$\int_R \phi_k(x)\phi_l(x)dx = \delta_{kl}$$

provided that the same quadrature factors are used to evaluate this integral as were used to reduce the integral equation (12) to matrix form as in (13) or (17).

The quadrature factors that are to be used in going from (12) to (13) or (17) need to be very carefully considered. First, even in the case of a single variable x, the choice of quadrature factors is by no means unique. The standard handbooks of mathematical formulae list the quadrature factors for the Trapezoid rule and Simpson's rule. The more complete work, Abramowitz and Stegun (1964), lists eight additional sets of quadrature factors, not to mention the open-type formulae which would be used to obtain quadrature factors like the $B_j^{(x_i)}$ of the preceding section. In the case of these formulae an error term is listed which generally is proportional to some derivative of the integrand evaluated in the interval of integration. This, at least, gives the impression that if the integrand is sufficiently "smooth", the higher the order of this derivative, the more exact the quadrature formula. All of these formulae are for equally-spaced abscissae, $x_{i+1}-x_i$=constant. When the abscissae are not equally spaced, one must be prepared to go to the basic relations from which such quadrature formulae are usually derived to obtain adequate expressions.

The situation when the parameter of integration is multidimensional is worse. Even the compendious Abramowitz and Stegun (1964) list only one quadrature formula for points located on a square grid and which are also on the boundary of the area over which one is integrating. It is usually mentioned that in dealing with a rectangular grid of points, one may make a multiple application of the one dimensional quadrature formulae. None of these cases describe the situation at hand, which we now explore to some extent.

To obtain the experimental values of the kernel $K(x_i,x_j)$ we have a fixed number of points, P_i, with coordinates (ξ_i,η_i) that are arbitrarily located. We cannot increase the number of points and we can do nothing about their location. The domain of integration, R, is determined by this network of points to a large extent, but not exactly. We may as well simplify the problem as much as possible by specifying that the boundary of R is a polygon obtained by connecting the points on the perimeter of this area. Note that the network of data points does not really define such a boundary uniquely. We need to specify something like a "convex" boundary before even the area of integration becomes uniquely defined. We do not consider all of the ramifications of the boundary selection any further since it would only make a difficult situation more difficult.

The ambiguity of the relation between the points, P_i, and the domain of integration is illustrated by the simple example of a re-entrant quadrilateral as shown in Figure IV-1. In A, the four points are considered to be bounded by a simple triangular region with the fourth point as an interior point and the region divided into three non-overlapping triangles. In B, the fourth point is considered to lie on the boundary of a re-entrant quadrilateral which is subdivided into two non-overlapping triangles.

Inside and on the boundary one has the points P_i. Now subdivide the region into elementary triangles with a point P_i at a vertex of each triangle. It is readily proved that if P = total number of points, B = number of boundary points, S = number of triangle sides, T = number of triangles, then

$$S = 3(P-1)-B$$
$$T = 2(P-1)-B$$

The fact that the boundary is not uniquely defined is reflected in the fact that B (which is included in the total number of points, P) is also a parameter that needs specification.

Consider now a single triangle that has for vertices the points 1, 2, 3. In order to carry out the integration of $f(\xi,\eta)$ over this triangle when only the values f_1, f_2, f_3 at the vertices are known, it is reasonable to consider a generalization (slight) of the trapezoid rule. Let the function be approximated by a plane over this triangle. It is easily shown that the integral over the triangle is given by

$$\int_{\Delta} f(\xi,\eta)\,d\xi d\eta = (f_1+f_2+f_3)\cdot(A/3) \tag{17}$$

where A is the area of the triangle.

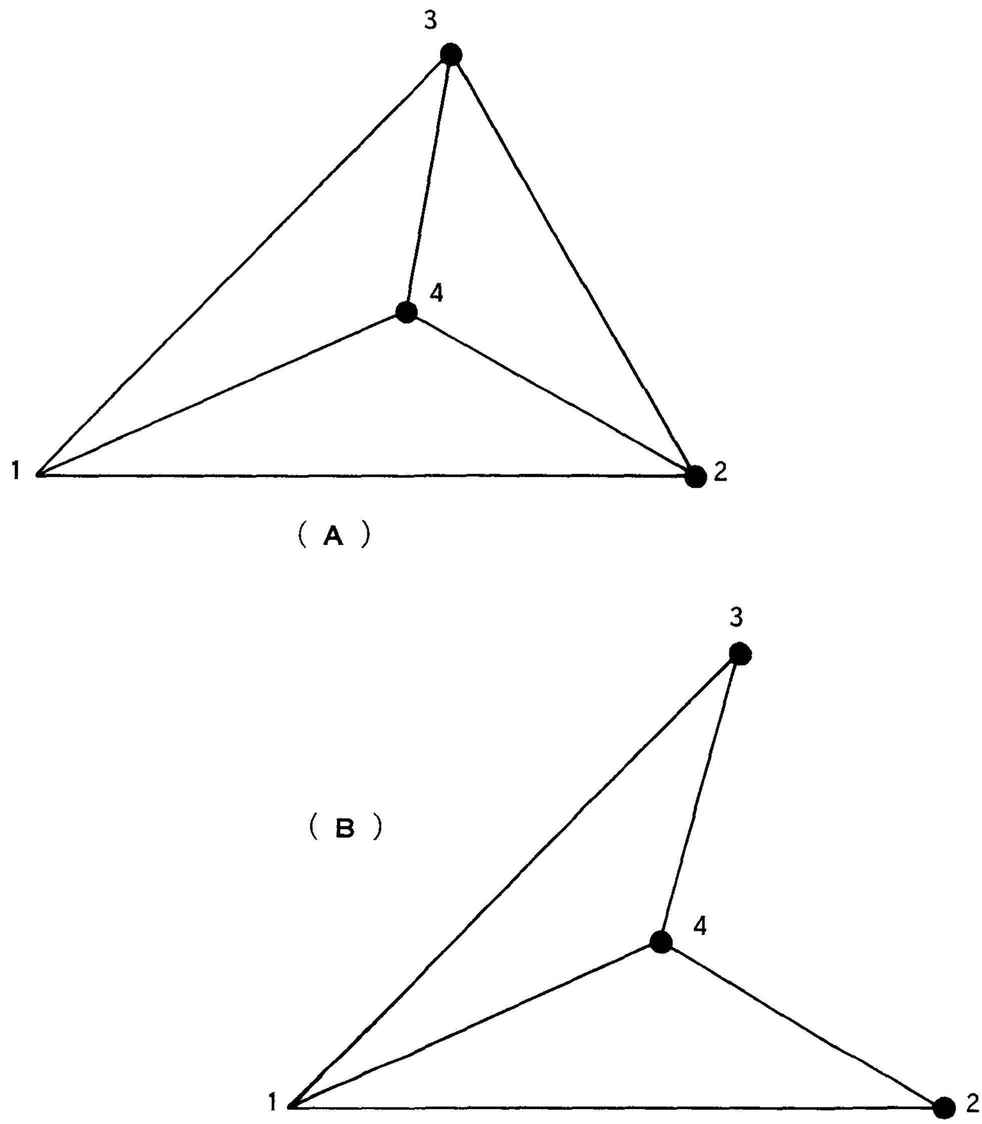


FIGURE IV-1

THE AMBIGUITY OF THE RELATION FOR THE NUMBER OF TRIANGLES AND SIDES DEPENDING ON BOUNDARY POINT ASSIGNMENT

Next consider the array of points that is subdivided into non-overlapping triangles that completely cover the region of integration. With each point, P_i, there will be one or more triangles with this point as a common vertex. The integral of $f(\xi,\eta)$ over the entire region R may then be approximated by the quadrature formula

$$\int_R f(\xi,\eta)\,d\xi d\eta \approx \sum_i f_i A_i$$

where i is the point index and A_i is given by

$$A_i = [\ \sum^* T_k(i)]/3$$

where $T_k(i)$ represents the area of the k'th triangle that has the point P_i as a vertex and $\sum_k^*$ is the sum of all such.

The situation is amenable to a simple geometric construction to visualize the areas thus associated with the points P_i. In an individual triangle, the lines joining the vertices with the mid-point of the opposite side divide the triangle into three quadrilaterals each with the same area. Thus, for the triangle 123, Figure IV-2, the points A,B,C are the midpoints of the sides and lines A1, B2, C3 all meet at the point Q, the "center of gravity" of the triangle. The quadrilaterals QB1C, QC2A, QA3B each have an area equal to 1/3 the area of the triangle 123.

When several triangles are put together to form a region subdivided into triangles, the area associated with each point is illustrated in Figure IV-3.

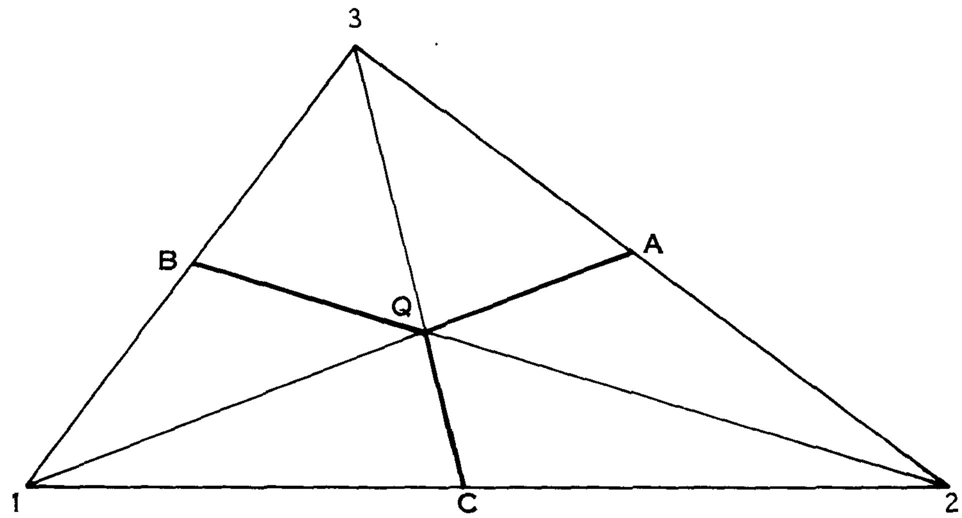


FIGURE IV–2

THE THREE EQUAL AREA QUADRILATERALS, 1CQB, 2AQC, AND 3BQA, RESULTING FROM JOINING THE VERTICES WITH THE MID – POINTS OF THE OPPOSITE SIDES OF THE TRIANGLE.

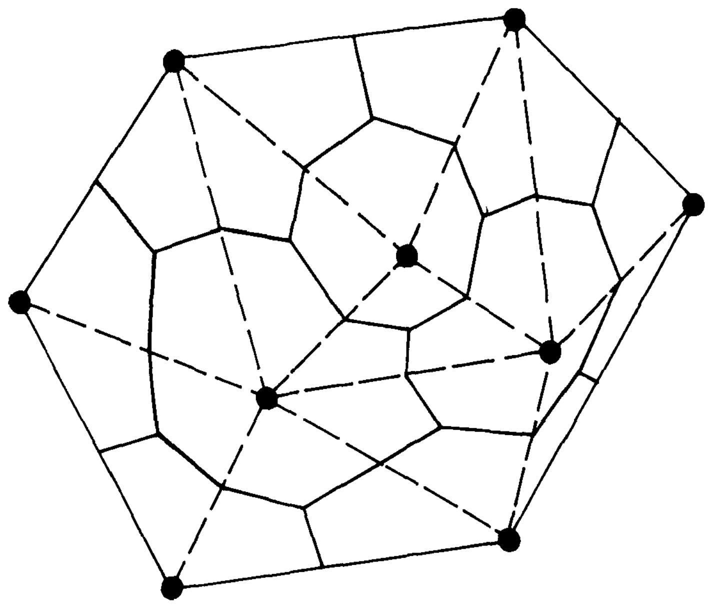


FIGURE IV—3

AN AREA SUBDIVIDED INTO TRIANGLES DETERMINED BY THE POINT LOCATIONS AND THE AREAS ASSIGNED TO EACH POINT ON THE BASES OF THE QUADRILATERALS IN EACH TRIANGLE ILLUSTRATED IN FIGURE IV—2.

The situation seems to be well under control at this point, but now consider how we divide a very simple convex quadrilateral into triangles. There are two choices of the way in which this is done, as shown in Figure IV-4. It is readily seen that the quadrature factors assigned to the points 1,2,3,4 in these two cases are vastly different from each other. When a region covered by many data points is considered, the number of ways that it may be divided into non-overlapping triangles becomes large, and each of these methods of subdivision will be associated with a different assignment of quadrature factors to the points. The problem then resolves itself into the question of what is the best way of subdividing the area covered by given points P_i into triangles so that the resulting quadrature formula will best approximate the integral concerned.

As an example of the wide variety of quadrature factors, consider a square with data points at $(\pm h, \pm h)$, $(\pm h, 0)$, $(0, \pm h)$, and $(0,0)$. There are four sub-squares and each can be divided into two triangles in two ways. There are thus 16 possible ways of dividing the area into non-overlapping triangles. Each of these will give a different system of quadrature factors based on the integration of a plane approximation over the triangles. These are listed in the following table. The quadrature factors in each case add to 24 so they are to be multiplied by $h^2/6$ where h is the spacing between points to get the correct units.

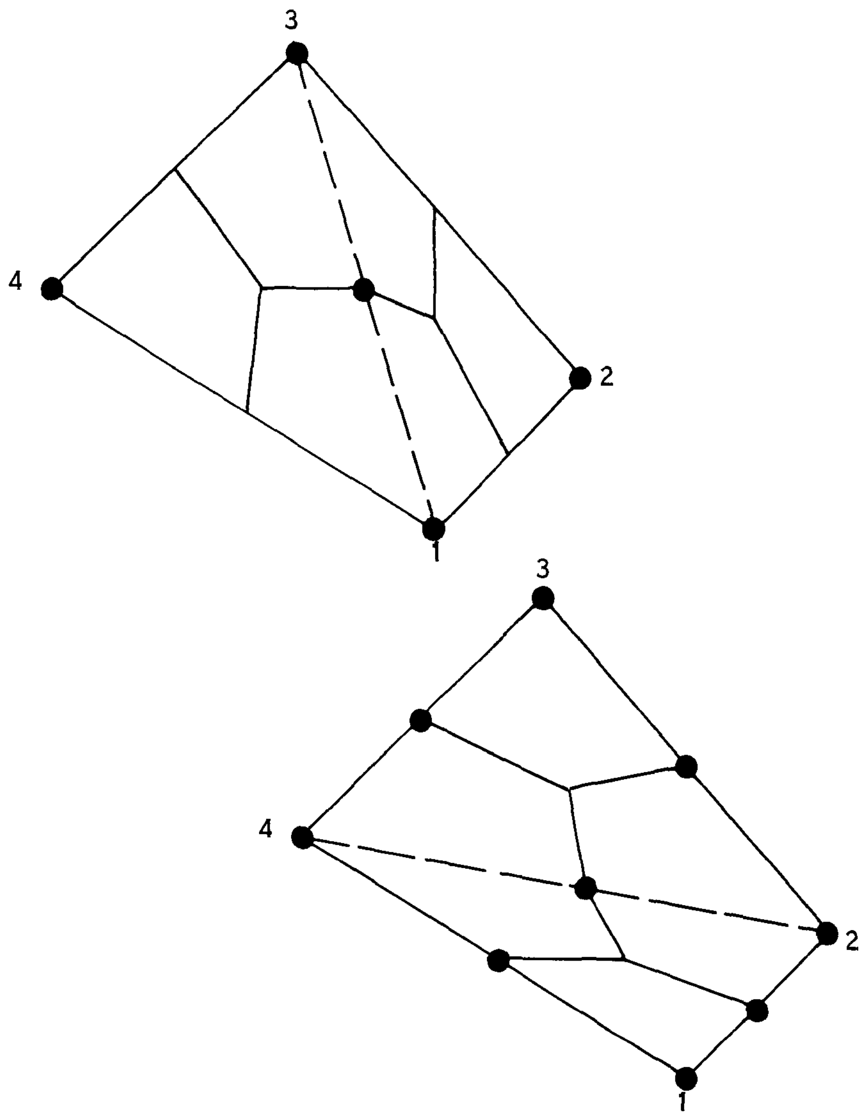


FIGURE IV–4

THE TWO WAYS THAT A CONVEX QUADRILATERAL MAY BE DIVIDED INTO TWO TRIANGLES AND THE RESULTING DIFFERENCE IN THE AREAS ASSIGNED TO EACH POINT.

Quadrature Factors for the 9 Points Covering a Square

Number	1	1	2	4	4	4	1 (Trap.)	1 (Simp.)
Coordinates								
(-h,+h)	1	2	1	1	1	1	3/2	2/3
(0,+h)	4	2	3	4	3	4	3	8/3
(+h,+h)	1	2	2	1	2	1	3/2	2/3
(-h,0)	4	2	3	3	3	3	3	8/3
(0,0)	4	8	6	6	7	5	6	32/3
(+h,0)	4	2	3	3	2	4	3	8/3
(-h,-h)	1	2	2	2	2	2	3/2	2/3
(0,-h)	4	2	3	2	2	3	3	8/3
(+h,-h)	1	2	1	2	2	1	3/2	2/3

The "number" row indicates the possible number of cases of each type: 1 is associated with a symmetrical arrangement, 2 indicates that a 90° rotation gives another arrangement but that a second rotation of 90° reproduces the original arrangement; 4 indicates that the original is not reproduced until the 4'th rotation through 90°.

The last two columns are symmetrical quadrature factors for this point array that are derived from the double application of the trapezoid rule and Simpson's rule respectively. The quadrature factors for the trapezoid rule (applied twice) are the average of those shown in the first two columns.

In this example, there appears to be no good over-all criteria for prefering one assignment of quadrature factors over another. In the case of each square, the two possible triangle subdivisions are symmetrical. One might prefer a symmetrical arrangement as in the first two columns, but this seems to be justified more on the basis of taste rather than a good substantiation of the better accuracy of the quadrature formula. When the points are not regularly arranged on a grid, but are more or less at random, one might prefer a subdivision into triangles that would favor a tendency toward equilateral triangles over long, skinny triangles, but again this appears to be justified more on the basis of taste than mathematics.

The standard hyperbolic paraboloid interpolation scheme (sometimes called double linear interpolation or iterated linear interpolation) over convex quadrilaterals might be used, but the same ambiguity on the resulting quadrature factors still exists because of the many ways that the observation net can be divided into convex quadrilaterals. In this case, one may be stuck with a few triangles since the quadrilateral subdivision need not "come out even".

c) <u>The Product Integral Technique</u>

If instead of interpolating the whole integrand on a triangle to obtain the quadrature factors one interpolates separately the two terms of the product, one obtains the relation

$$I_\Delta \cong \int_\Delta K(x,y;x',y')\phi(x',y')dx'dy' = 2A\int_0^1 \int_0^{1-q} K(x,y;p,q)\phi(p,q)\,dp\,dq$$

where

$$\phi(p,q) = \phi_1 + p(\phi_2-\phi_1) + q(\phi_3-\phi_1)$$

$$K(x,y;p,q) = K_1 + p(K_2-K_1) + q(K_3-K_1)$$

and in which ϕ_1, ϕ_2, ϕ_3 and K_1, K_2, K_3 are the values of the proper function and the kernel function at the triangle corners and A is the triangle area. The results of this integration lead to

$$I_\Delta \cong (A/12)[\phi_1(2K_1+K_2+K_3)+\phi_2(K_1+2K_2+K_3)+\phi_3(K_1+K_2+2K_3)]$$

which may be expressed in matrix/vector form as

$$I_\Delta \cong (A/12)\{K_1, K_2, K_3\} \begin{Bmatrix} 2, & 1, & 1 \\ 1, & 2, & 1 \\ 1, & 1, & 2 \end{Bmatrix} \begin{Bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{Bmatrix}$$

where now one has a matrix of quadrature factors. Sum over all of the triangles that cover the region of integration to obtain the quadrature form for the integral equation as

$$\lambda\phi_i = \sum_j [\sum_k K_{ik}A_{kj}]\phi_j = \sum_k K_{ik}[\sum_j A_{kj}\phi_j] \qquad (18)$$

where $\phi_i=\phi(x_i, y_i)$, $K_{i,k} = K(x_i, y_i; x_k, y_k)$, and the elements of the matrix of quadrature factors, $\{A_{ij}\}$ are defined as

$$A_{ii} = (2/12)\sum(\text{Areas of } \Delta\text{'s with } P_i \text{ as common vertex})$$

$$A_{ij} = (1/12)\sum(\text{Areas of } \Delta\text{'s with } P_iP_j \text{ as common side})$$

Write the above in matrix/vector form as

$$\lambda\phi = KA\phi$$

where ϕ is the column vector of proper values $\phi(x_i)$. Now assume for the moment that A is a positive definite matrix. (It is symmetric.) It will then have a square root which we will denote as $A^{1/2}$. Multiply on each side by $A^{1/2}$ to obtain

$$\lambda(A^{1/2}\phi) = (A^{1/2}KA^{1/2})(A^{1/2}\phi).$$

If, now, we let $A^{1/2}\phi=\theta$, we have the relation

$$\lambda\theta = (A^{1/2}KA^{1/2})\theta$$

which is the analogue of the relation (17) obtained when the quadrature factors were simple scalars associated with each point.

The solutions of the above matrix problem, $\theta_k(x_i), \lambda_k$, are such that, as before,

$$\sum_i \theta_k(x_i)\theta_l(x_i) = \delta_{kl}.$$

Then in terms of the values $\phi_k(x_i)$, one has

$$\theta_k(x_i) = \sum_m A_{im}^{1/2}\phi_k(x_m)$$

so that

$$\sum_i \theta_k(x_i)\theta_l(x_i) = \sum_i (\sum_m A_{im}^{1/2}\phi_k(x_m))(\sum_n A_{in}^{1/2}\phi_l(x_n))$$

$$= \sum_m \sum_n \phi_k(x_m)\phi_l(x_n)(\sum_i A_{im}^{1/2} A_{in}^{1/2})$$

where, in the above, $\{A_{ij}^{1/2}\}$ stands for the elements of $A^{1/2}$, the square root matrix of the matrix A (<u>not</u> the square roots of the elements of A). Since $A^{1/2}$ is symmetric (as was A) the summation on i in the above simply gives the elements of A. Thus

$$\sum_i \theta_k(x_i)\theta_l(x_i) = \sum_m \sum_n A_{mn}\phi_k(x_m)\phi_l(x_n).$$

This is the same result that would have been obtained had we evaluated the "product integral" expression

$$\int_R \phi_k(x)\phi_l(x)dx = \delta_{kl}$$

using the same individual interpolation formula for the factors in the integrand that was used to evaluate the product integral that appears in the initial integral equation.

The square root matrix of the matrix A is readily obtained by recourse to the proper functions and values of the matrix A. If the proper values/functions of A are μ_k, ψ_{ik} so that M is the diagonal matrix of the μ_k's and Ψ the matrix of which ψ_{ik} are the column vectors, then $A = \Psi M \Psi'$. Since for a symmetric positive definite matrix the proper values are real and positive, then the proper values have square roots (use the positive sign). Then the matrix $\Psi M^{1/2}\Psi'$, where $M^{1/2}$ is the diagonal matrix of elements $\mu_k^{1/2}$ on the principal diagonal and zeros elsewhere, is the square root matrix of A, i.e., $A^{1/2} = \Psi M^{1/2}\Psi'$. Note that $A^{1/2}A^{1/2} = (\Psi M^{1/2}\Psi')(\Psi M^{1/2}\Psi') = \Psi M^{1/2} I M^{1/2}\Psi' = \Psi M^{1/2}M^{1/2}\Psi' = \Psi M \Psi' = A$.

d) Test for the Accuracy of Solutions

In view of the fact that the quadrature factors for the numerical solution of an homogeneous Fredholm integral equation of the second kind are not well-defined quantities, a test of the validity of solutions seems to be in order. The usual mathematical error estimates are unsatisfactory for this since they all require more information than is available from experimental data on the kernel function. A test is available, however; namely a simple comparison of the proper functions arising from two equally valid choices for the quadrature factors. Thus, let there be two choices of quadrature factors, A_i and B_i. The corresponding algebraic equations to be solved are

$$\lambda[\phi(x_i)\sqrt{A_i}] = \sum_j [\sqrt{A_i}K(x_i,x_j)\sqrt{A_j}][\sqrt{A_j}\phi(x_j)]$$

$$\lambda[\phi(x_i)\sqrt{B_i}] = \sum_j [\sqrt{B_i}K(x_i,x_j)\sqrt{B_j}][\sqrt{B_j}\ \ (x_j)]$$

and let

$$\theta_k^{\#}(x_i) = \sqrt{A_i}\phi_k(x_i)$$

$$\theta_k^{*}(x_i) = \sqrt{B_i}\phi_k(x_i)$$

be the proper functions at the points x_i and let $\lambda_k^{\#}$, λ_k^{*} be the corresponding proper values. Then consider the sum of the squares of the differences of these solutions. If the solutions are reasonably similar, this quantity should be very close to zero. On expanding the square and summing over the points x_i, one obtains

$$\sum_i[\theta_k^{\#}(x_k)-\theta_k^{*}(x_i)]^2 = \sum_i[\theta_k^{*}(x_i)]^2-2\sum_i \theta_k^{*}(x_i)\theta_k^{\#}(x_i) + \sum_i[\theta_k^{\#}(x_i)]^2.$$

The first and last term on the right are each unity so that

$$\sum_i[\theta_k^{*}(x_i)-\theta_k^{\#}(x_i)]^2 = 2[1-\sum_i \theta_k^{*}(x_i)\theta_k^{\#}(x_i)].$$

The second term in brackets on the right is simply the "correlation coefficient" for the point by point comparison of the two solutions. (Each has unit second moment about zero, but need not have a first moment of zero, although for proper functions of order greater than a small number this is essentially the case. We compute the sum of products as indicated ignoring the fact that the first moment may not be zero. It is an immediate consequence of the Schwartz inequality that the value of this sum of products lies between +1 and -1 so it "looks like" a correlation coefficient anyway.)

The quantity

$$C_k = \sum_i \theta_k^*(x_i)\,\theta_k^\#(x_i)$$

serves as a test parameter for the "self consistency" of the two solutions $\theta_k^*(x_i)$ and $\theta_k^\#(x_i)$. If C_k is close to 1, it may be said that the solutions are self consistent. If it departs from 1 by a significant amount, then the solutions are not self consistent. A better condition for self consistency is much stronger than this. We have a sequence of values C_k, k=1, ..., n. It is expected that for k small, the values of the test parameter will be close to one. If at some value of k, say k^*, the value of C_{k*} has decreased abruptly compared to the previous values, then we can say that not more than k*-1 solutions are self consistent. This is regardless of whether or not the subsequent values of C_k, k>k*, may be close to 1. This is because in each sequence of solutions every solution is orthogonal to all previous solutions of the sequence.

One cannot say which of the solutions is incorrect if $C_k<1$ when the quadrature factors are equally valid. One can only say that k*-1 solutions appear to be equally valid and that any further solutions in the sequence are suspect.

From the preceding, it is concluded that by formulating the problem of matrix reduction (outlined in Sub-section 1 and treated from the Factor Analysis point of view in Sub-section 2 as an Integral Equation) there are ambiguities of technique that have an important bearing on the accuracy of determining the proper functions. These ambiguities may be used to test for self-consistent solutions. If the solutions are self-consistent one can say nothing about their accuracy. Accurate solutions must be self-consistent, but inaccurate ones may be also. On the other hand, if a pair of solutions are inconsistent, then one or the other or both must be inaccurate. This has an important bearing on the value of k determined from the Factor Analysis method (k = number of statistically significant solutions to the matrix problem) as outlined in Section 2. It seems reasonable that if there are not more than k* solutions of the integral equation formulation that can be reasonably accurate from a mathematical point of view, then using more than k* solutions to the matrix problem appears to be of doubtful value, even if the appropriate statistical tests indicate that they are "statistically" significant ($k^* < k$).

C. ANALYSIS OF ST. LOUIS SO_2 DATA

1. The Basic Data

The daily data on SO_2 concentrations received from the EPA were analyzed by means of the proper function/value method to determine the amount of the residual variances (Ruff, 1973a, 1973b). The data cover 89 days in 1964-1965 and were for 40 locations shown in Figure IV-5. (The triangle indicates the location of dual 2-hour and 24-hour samples). Not all stations had records that permitted the assignment of an SO_2 concentration value to each day. It was found that if only those days were used in which data were available from all stations, there were only 27 such. This meant that the

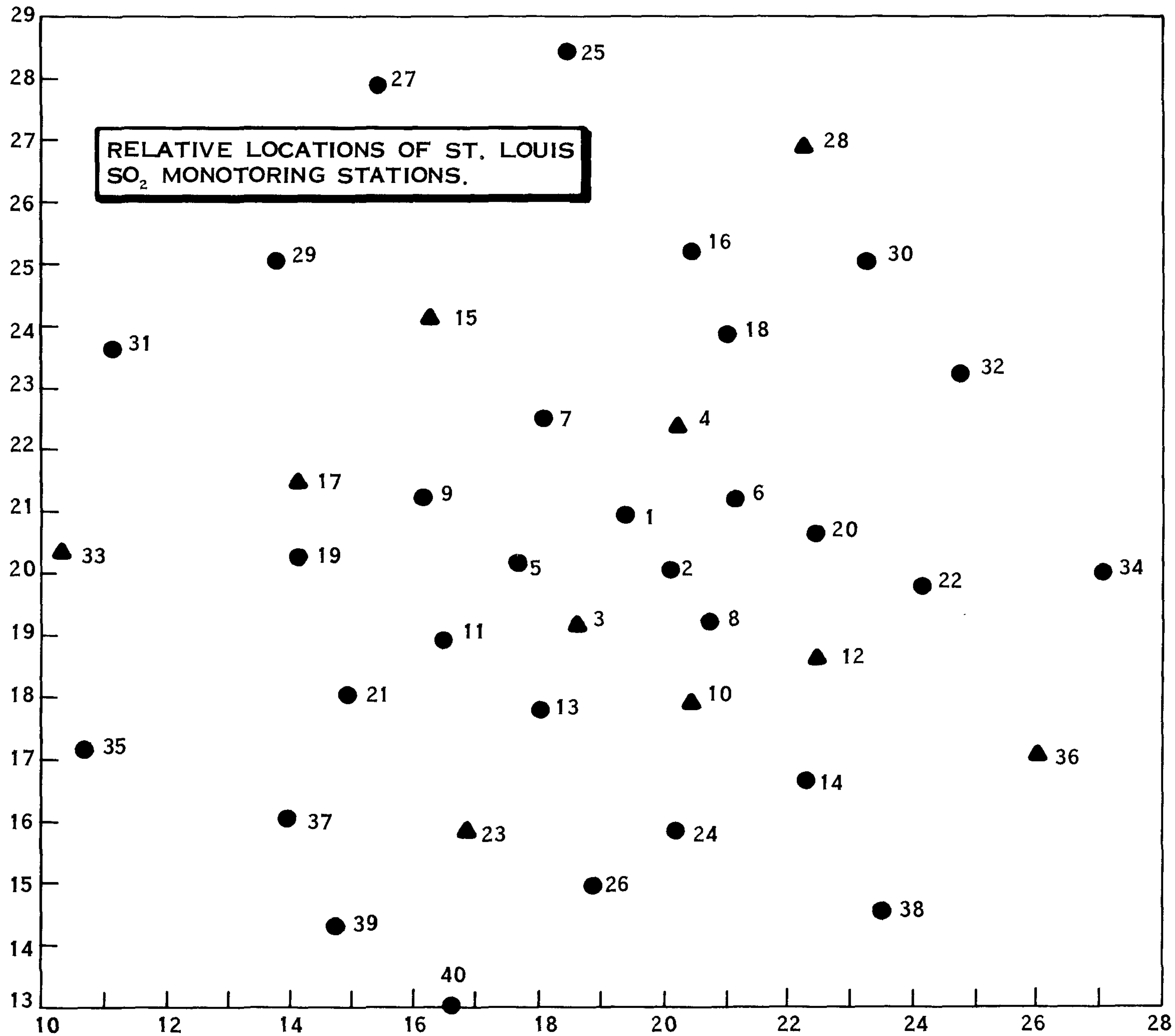


FIGURE IV—5

covariance matrix would be of rank 27 instead of a more desirable rank of 40 (see Appendix B). Inspection of the data for missing days by stations revealed that there were a large number of days on which data for only one or two stations were missing. It was felt that it would be desirable to "interpolate" data for the missing stations under such circumstances. This would increase the validity of the data as a whole without seriously degrading that for the single station at which the interpolated data was inserted. The result of this synthetic increase in the data base resulted in an increase to 59 effective data days, which in turn assured a full rank of 40 for the covariance matrix.

The area covered by the stations was divided into two systems of non-overlapping triangles covering a region with common exterior boundary in each case. The station coordinates and the triangle assignments are listed in Tables IV-1 and IV-2. It will be noted in Table IV-1 that though the areas assigned to points are correlated, there are also some large variations between the areas assigned to a given point, the ratio in some instances being larger than 2:1. The arrangements of the triangles are also shown in Figures IV-6 and IV-7. The station coordinates are in grid units as in Ruff (1973a). The areas assigned to the points in Table IV-1 are in square grid units and represent one third of the area of the triangles which have the point concerned as a common vertex as on page 79.

The triangle assignments were made in order to obtain quadrature factors as outlined in Sections 1b and 1c. Only the points shown in Figure IV-5 with coordinates given in Table IV-1 at which covariance data are available are given a priori. To obtain quadrature factors which are associated with each of these points (that is, the areas that have been tabulated in Table IV-1 under the columns headed 1 and 2

Table IV-1. Station Coordinate and Area Assigments

Station Index	Coordinates		Area	
	X	Y	1	2
1	19.42	20.86	1.77	3.20
2	20.16	20.06	2.00	1.15
3	18.66	19.16	1.50	2.89
4	20.24	22.36	3.53	3.50
5	17.72	20.14	3.35	1.76
6	21.18	21.20	2.48	3.16
7	18.12	22.50	4.70	5.04
8	20.76	19.20	3.89	2.46
9	16.22	21.16	3.65	4.96
10	20.48	17.88	1.53	3.60
11	16.52	18.92	3.22	3.40
12	22.48	18.64	1.59	4.52
13	18.04	17.76	5.68	4.48
14	22.32	16.64	8.63	4.87
15	16.32	24.12	11.14	10.85
16	20.42	25.20	9.87	7.77
17	14.14	21.46	9.19	7.29
18	21.02	23.84	3.88	3.94
19	14.16	20.26	4.97	6.39
20	22.44	20.64	6.50	2.83
21	14.96	18.04	4.80	6.20
22	24.14	19.78	3.62	5.24
23	16.88	15.88	4.48	6.34
24	20.24	15.80	3.42	4.77
25	18.50	28.42	4.20	6.10
26	18.88	14.92	5.90	3.88
27	15.42	27.88	6.01	3.46
28	22.30	26.86	0.90	2.38
29	13.78	25.06	2.47	4.59
30	23.26	25.04	3.20	2.44
31	11.14	23.64	5.08	3.61
32	24.76	23.22	5.71	5.43
33	10.34	20.34	4.01	4.71
34	27.10	19.98	5.96	3.56
35	10.68	17.18	5.10	4.97
36	26.04	17.06	3.14	5.15
37	13.96	16.06	3.51	3.09
38	23.54	14.52	4.70	4.70
39	14.74	14.32	3.75	1.77
40	16.64	13.04	2.71	3.57

Table IV-2. Triangle Assignments

Triangle Index	No. 2			No. 1			Triangle Index	No. 2			No. 1		
	1	2	3	1	2	3		1	2	3	1	2	3
1	26	38	40	26	38	40	31	10	13	14	23	24	26
2	24	26	38	24	26	38	32	8	10	14	13	23	24
3	14	24	38	14	24	38	33	8	12	14	10	13	24
4	14	36	38	14	36	38	34	12	14	22	10	14	24
5	14	34	36	12	14	36	35	8	12	22	10	12	14
6	14	22	34	12	22	36	36	8	20	22	8	10	12
7	20	22	34	22	34	36	37	2	8	20	8	12	20
8	20	32	34	22	32	34	38	2	6	20	12	20	22
9	18	20	32	20	22	34	39	6	18	20	6	8	20
10	16	18	32	6	20	32	40	4	6	18	2	6	8
11	16	30	32	4	6	32	41	4	16	18	1	2	6
12	16	25	30	4	18	32	42	4	15	16	1	4	6
13	25	28	30	18	30	32	43	4	7	15	1	4	7
14	16	25	27	16	18	30	44	7	15	17	4	7	18
15	15	16	27	16	28	30	45	7	9	17	7	16	18
16	15	27	29	16	25	28	46	9	17	19	7	15	16
17	15	29	31	15	16	25	47	9	19	21	7	9	15
18	15	17	31	15	25	27	48	9	11	21	9	15	17
19	17	31	33	15	27	29	49	11	21	23	9	17	19
20	17	33	35	15	17	29	50	11	13	23	9	11	19
21	17	19	35	17	29	31	51	5	11	13	11	19	21
22	19	35	37	17	31	33	52	3	5	13	11	13	21
23	19	21	37	17	19	33	53	3	8	13	13	21	23
24	21	37	39	19	33	35	54	8	10	13	3	11	13
25	21	23	39	19	21	35	55	2	3	8	3	10	13
26	23	26	39	21	35	37	56	2	3	5	3	8	10
27	13	23	26	21	23	37	57	1	2	5	2	3	8
28	26	39	40	23	37	39	58	1	2	6	1	2	3
29	13	24	26	23	39	40	59	1	6	7	1	3	5
30	13	14	24	23	26	40	60	4	6	7	3	5	11
							61	1	5	7	5	9	11
							62	5	7	9	1	5	9
							63	5	9	11	1	7	9

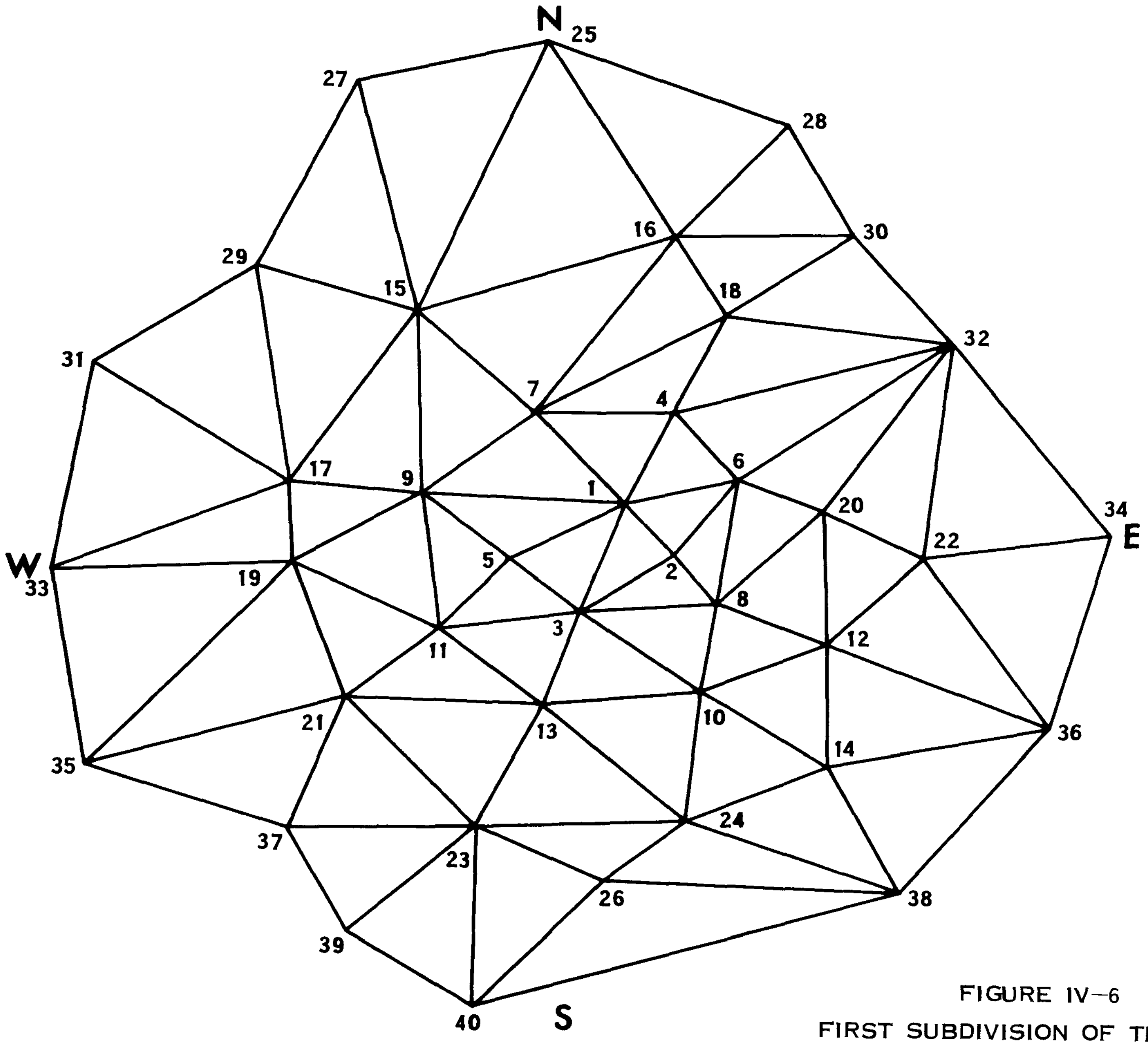


FIGURE IV–6

FIRST SUBDIVISION OF THE AREA COVERED BY 40 SO_2 MEASURING POINTS AT ST. LOUIS.

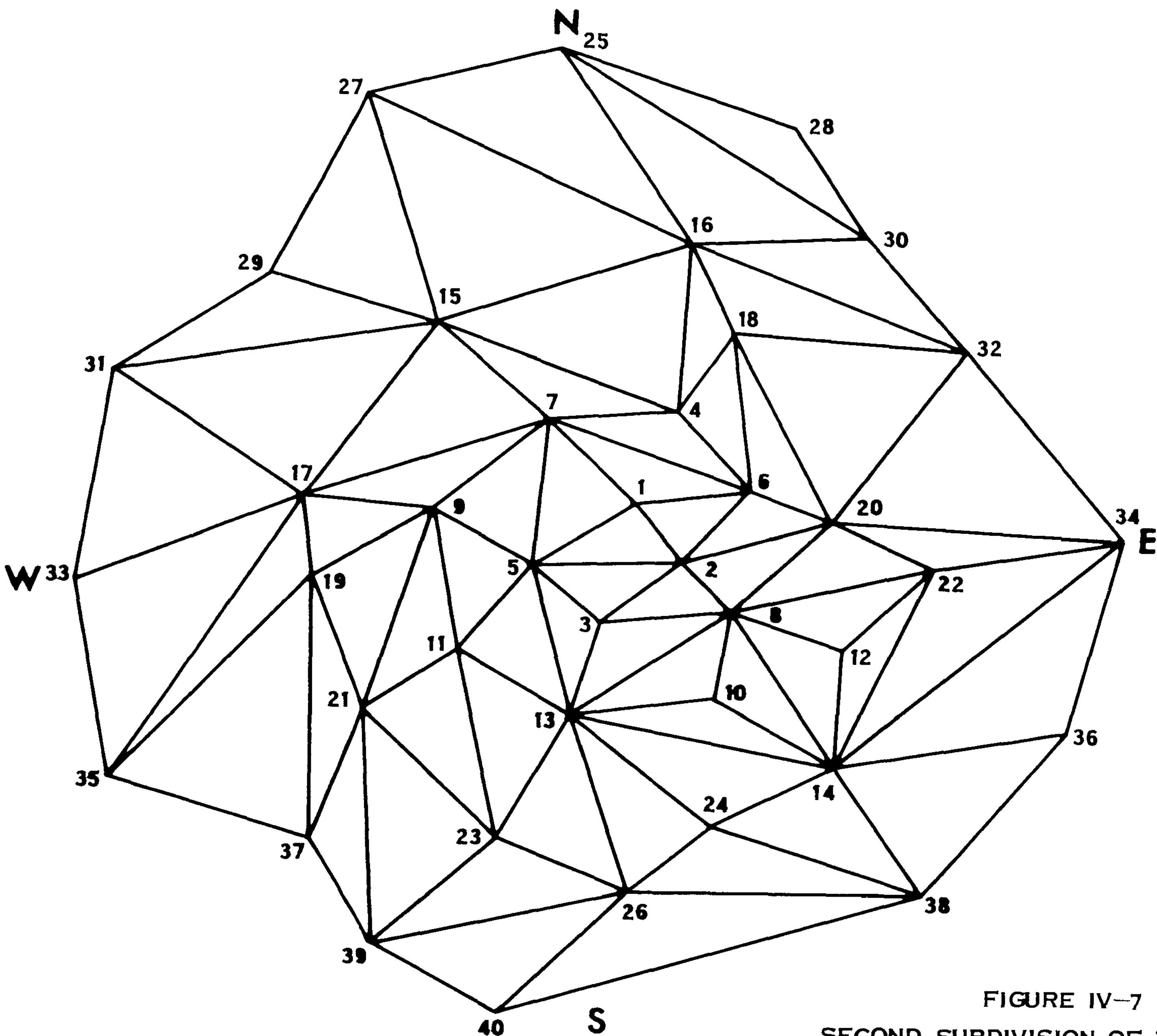


FIGURE IV–7

SECOND SUBDIVISION OF THE AREA COVERED BY 40 SO_2 MEASURING POINTS AT ST. LOUIS.

since we are to compare two different ways of obtaining the quadrature factors or areas) the entire region is covered by non-overlapping triangles and to each point is assigned the area equal to one third of the area of all triangles with a vertex at this point. This covering of the region by triangles is not unique, so we say that we make triangle "assignment" or "assignments" since there is some freedom of choice here. For the purpose at hand, only two such "assignments" are made. Many others could have been made.

The areas, or quadrature factors, for each of the two assignments of the triangle coverage (Table IV-1) that are associated with each point are "correlated" in the sense that if for one assignment of triangle coverage the quadrature factor (or area) associated with a given point is small (large) then for another assignment of triangle coverage it will also tend to be small (or large). In other words, if one were to compute the product moment correlation coefficient for the areas (quadrature factors) shown in columns 1 and 2 under Area in Table IV-1, this correlation coefficient would be positive and significantly different from zero. This is simply due to the fact that, as shown in Figure IV-5, in some areas of the region the points are more dense than in others. The statement that the areas (or quadrature factors) are correlated would be true for any two ways of covering the area with non-overlapping triangles, not just the two that happen to have been selected here.

2. Quadrature Factors and Techniques

In order to obtain a comparison between the proper values/functions for the different area assignments (triangle coverages) these were computed and compared using the "trapezoid method" and the "product integral technique". The "trapezoid method" refers to the assignment of quadrature factors on the basis of fitting the entire integrand as a plane over each triangle. The "product integral technique" refers to fitting each factor in the integrand separately as a plane over each triangle.

With reference to the quadrature factors used in the "trapezoid method", a search was made of the literature on the subject. It was found that the texts Davis and Rabinowitz (1967) and Stroud (1971) were devoted to more mathematical aspects of the quadrature problem in several dimensions. This was true also of papers found in the recent literature such as Ewing (1941), Tyler (1953), Synge (1953), Mises (1954), Thacher (1957), Hammer and Wymore (1957), Hammer and Stroud (1958), Albrecht and Collatz (1958), Stroud (1960a and 1960b), Ceschino and Letin (1972) to mention only a few. Even Dixon (1973), a survey paper, did not list multidimensional quadrature formulas that could be applied. Almost invariably mathematical interest has been confined to quadrature formulas wherein the function values are given at arrays of points prescribed in advance. These are, of course, useless when the data points are given in an essentially random manner. It was found that Mises (1936) did give a formula that could be applied to an arbitrarily given triangle. It is for this reason that in Section 1b the derivation of the quadrature factors for the "trapezoid method" was developed in some detail from "first principles" and the implications of the use of such a formula (especially the ambiguity of the triangle selections) were pointed out. We have not seen this kind of treatment of the problem in any published work. (On the other hand, it is so elementary that we feel sure that it must have been treated before somewhere and that our search has not been sufficiently exhaustive to find it.)

With respect to the "product integral technique", it was found after the method of Section 1c had been developed that a somewhat similar procedure had been developed by Boland and Duris (1969 and 1971) and Boland (1972) but, again, only for the case in which the function is evaluated at a prescribed regular array of points in one dimension. See also Beard (1947) for an earlier treatment, also in one dimension. We have not found anything treating the case at hand, arbitrarily preassigned points in two dimensions.

The proper values so obtained are shown in Figure IV-8 for the trapezoid method and Figure IV-9 for the product integral method. Each figure shows the proper values for each triangle assignment for index values 20-40, one by a triangle and the other by a dot which appear one above the other. For the index values 1-19 the differences were so small that the points are scarcely distinguishable and consequently only the dots are shown. The differences in the assignments of the areas to each data point apparently make little difference in the proper values obtained. On the other hand, comparison between figures shows at once that for index 5 through 40 the proper values obtained by using the product integral method of evaluating the integral yields proper values that are distinctly less than those obtained by the trapezoid method. The first ten proper values are compared in more detail in Table IV-3 in which the ratio decreases somewhat irregularly from 0.9 at index 1 to near 0.5 at index 5 and higher. This is discussed in more detail on p. 104.

The "knee" in the curve of log proper value vs index at index 5 or 6 represents an interesting phenomena which has been noted by Craddock and Flintoff (1970) and investigated

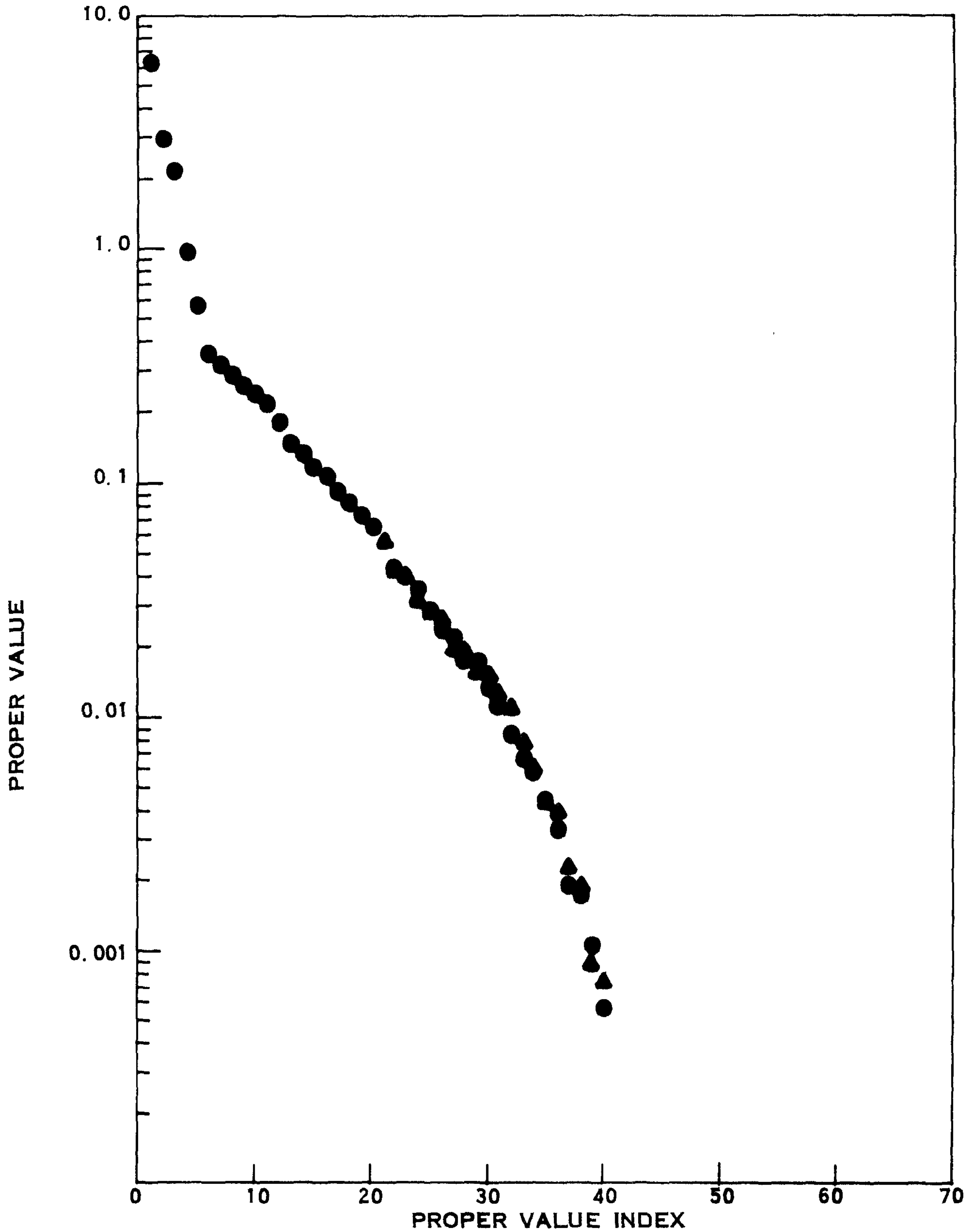


FIGURE IV-8

PROPER VALUES FOR DIFFERENT TRIANGLE ASSIGNMENTS USING THE TRAPEZOID METHOD.

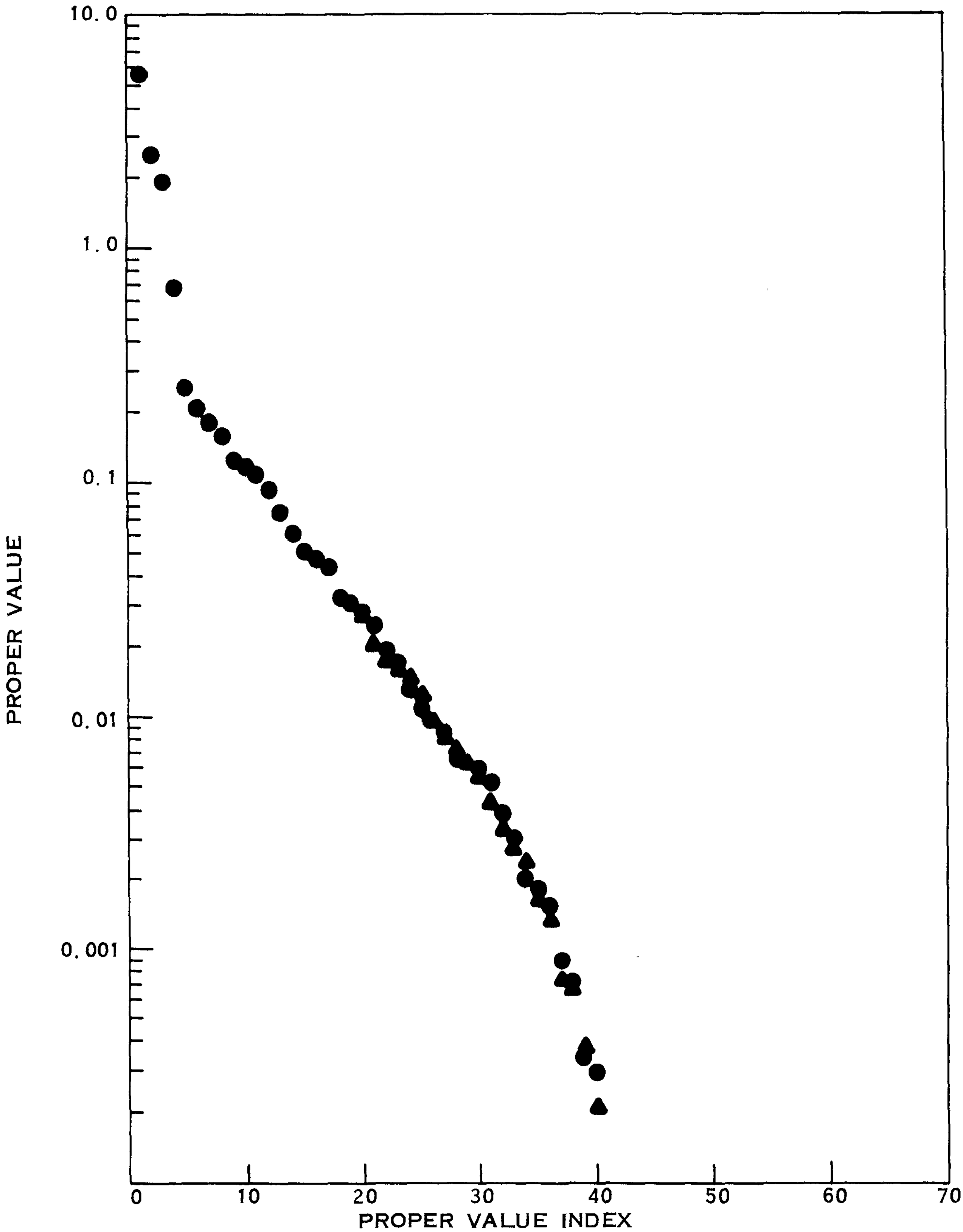


FIGURE IV—9

PROPER VALUES FOR THE DIFFERENT TRIANGLE ASSIGNMENTS USING THE PRODUCT INTEGRAL METHOD

Table IV-3. Comparison of the First Ten Proper Values by the Different Computing Methods

	Proper Values		
Index	Trapezoid	Product Integral	Ratio
1	6.28810	5.71721	0.90921
2	3.02629	2.49737	0.82522
3	2.23924	1.88175	0.84035
4	0.98689	0.68170	0.69075
5	0.57566	0.24709	0.42923
6	0.35671	0.18841	0.52819
7	0.32275	0.16377	0.50742
8	0.29378	0.16263	0.55358
9	0.26159	0.11669	0.44608
10	0.23855	0.11485	0.48145

synthetically by Farmer (1971). Craddock and Flintoff remark to the effect that this marks the point beyond which the proper functions appear to be more or less random, but they give no firm criteria on which they base their measure of randomness. We will be able to provide a criterion that seems to fit the situation rather well.

In order to compare the proper functions obtained by the two triangle subdivisions and the two quadrature techniques the "self consistency" test parameter, C_k, i.e., the correlation of the proper functions for the methods being compared, was computed. The values of $|C_k|$ as a function of index number are illustrated in Figures IV-10 and IV-11. The absolute value is used since there is always a basic ambiguity in the sign of the proper functions.* In Figure IV-10 it is to be noted that the first significant drop in the self consistency test parameter occurs after the 7'th index where the quadrature by the trapezoid method is used and after the 6'th index where quadrature is by the product integral method. These are one index higher than the knee in the curves of log proper value against index. Although some of the test parameter values are reasonably large for higher values of the index, it does not seem prudent to admit validity to the corresponding proper functions. Each proper function is orthogonal to those of lower index and low values of the self consistency criterion precede these larger self consistency criterion values.

---

* The proper functions are the solutions of the integral equation (12), page 70 (or of its algebraic counterpart when formulated in discrete terms). This integral equation is "homogeneous". It is readily seen that if $\phi(x)$ is a solution, then $-\phi(x)$ is also a solution. When solved in discrete terms, a standard eigenvalue/eigenfunction computation routine may be used. Whether one obtains $\phi(x)$ or $-\phi(x)$ from such a computation routine is more or less a matter of chance. Either one is valid.

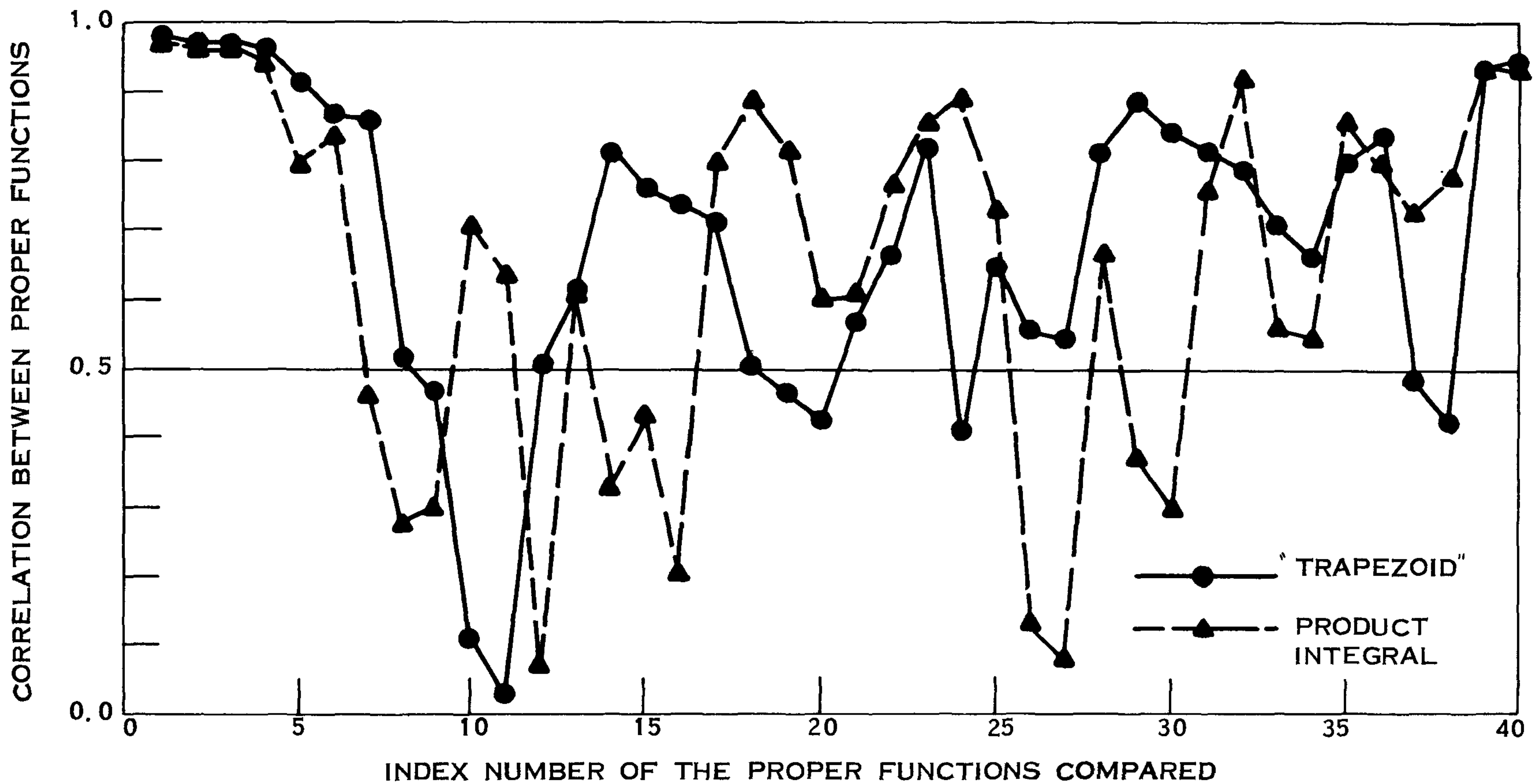


FIGURE IV–10

CORRELATION BETWEEN PROPER FUNCTIONS FOR DIFFERENT TRIANGLE ASSIGNMENTS

The self consistency criterion between the different quadrature methods for the two triangle subdivisions is illustrated in Figure IV-11. The situation seems to be similar to that of the comparison between triangle subdivisions of the preceding figure except that the first "break" in the test parameter appears to occur following the fourth index. It is to be noted that this is the point at which the ratio of the proper values obtained took a very abrupt drop to below 0.5 (and remained in the neighborhood of 0.5 thereafter).

The reasons for both the abrupt drop in the ratio of proper values for the two different computing procedures (trapezoid vs product-integral methods) and the drop in the correlation between proper functions for these two computing procedures lies in the fact that the product-integral technique is in effect a procedure which "smooths" the kernel function as compared with the trapezoid method. Thus, from (13), page 71, for the trapezoid method we have (slight obvious change in notation)

$$\lambda\phi_i = \sum_j K_{ij}A_j\phi_j$$

while from (18), page 86, for the product-integral technique

$$\lambda\phi_i = \sum_j \left(\sum_k K_{ik}A_{kj}\right)\phi_j$$

where A_j is defined on page 79 and A_{kj} on page 86. In order that the relations above be equivalent, it needs to be shown that

$$\sum_k K_{ik}A_{kj} = \tilde{K}_{ij}A_j$$

where A_j and A_{kj} are defined as above and K^*_{ij} is the required smoothed value. In other words, it will be necessary to show that

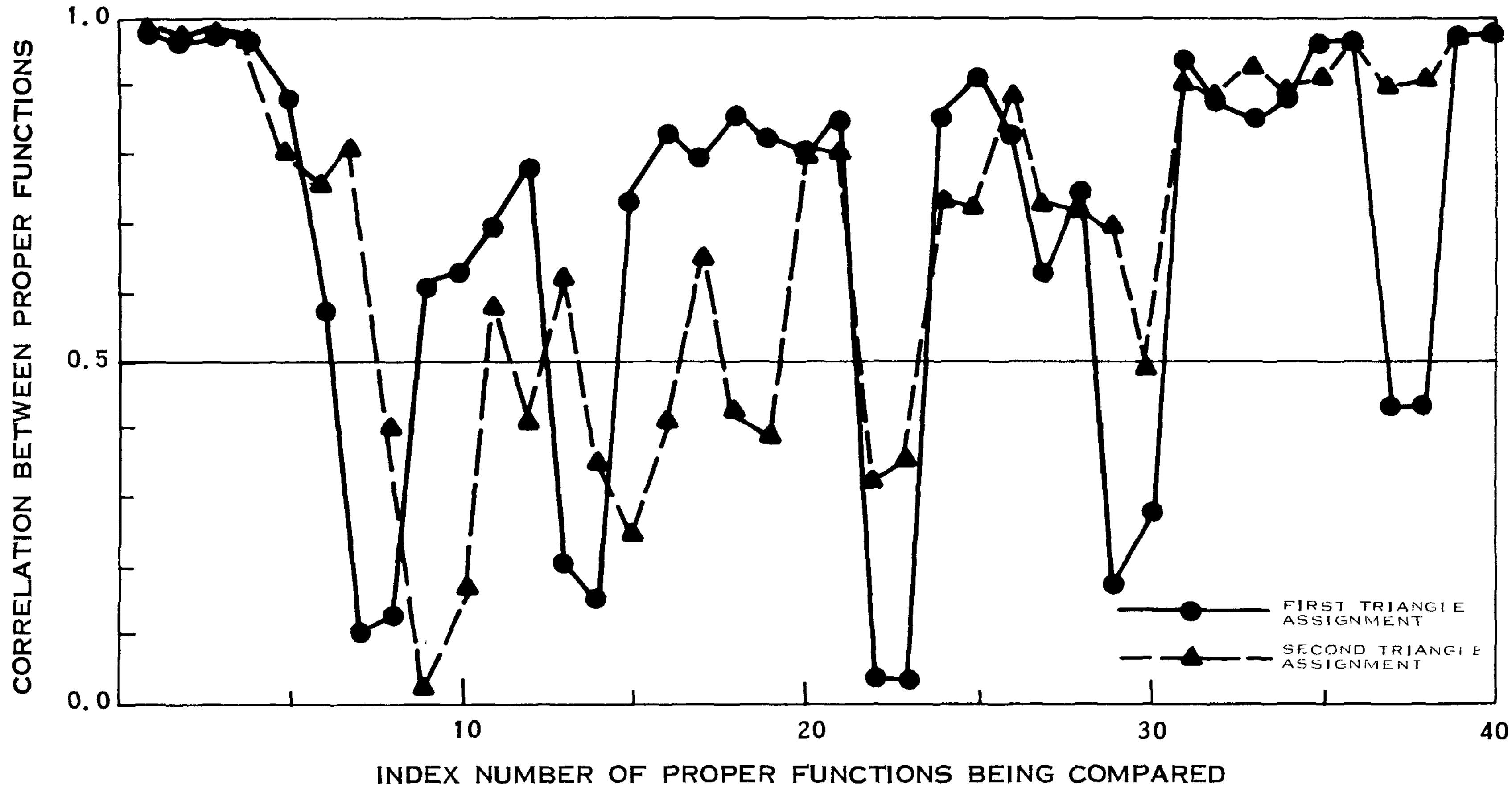
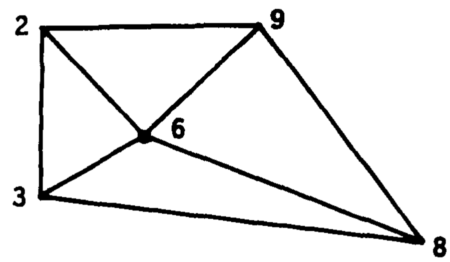


FIGURE IV–11

CORRELATION BETWEEN PROPER FUNCTIONS USING DIFFERENT QUADRATURE TECHNIQUES. REPRESENTED ARE THE TWO DIFFERENT BUT APPARENTLY EQUALLY VALID TRIANGLE ASSIGNMENTS.

$$\sum_k A_{kj} = A_j$$

so that K^*_{ij} will be the weighted average of the values K_{ik}. Rather than go through the details of a formal demonstration in the abstract, we consider only a concrete example which illustrates what is involved. In the diagram (no number) consider the point j and for the example we take j=6. Let the triangle selection be such that the triangles with 6 as the common vertex be (2,3,6), (3,8,6) (8,9,6), (9,2,6).



Since $A_{kj}=0$ unless k=j=6 or the edge P_kP_6 is involved, there are only five values that are not zero; A_{26}, A_{36}, A_{66}, A_{86}, A_{96} and these have the values as defined on page 86:

$$A_{26} = [(9,2,6) + (2,3,6)]/12$$
$$A_{36} = [(2,3,6) + (3,8,6)]/12$$
$$A_{66} = 2[(2,3,6) + (3,8,6) + (8,9,6) + (9,2,6)]/12$$
$$A_{86} = [(3,8,6) + (8,9,6)]12$$
$$A_{96} = [(8,9,6) + (9,2,6)]/12$$

where (i,j,k) indicates the area of the triangle with vertices P_i, P_j, P_k and (i,j,k) must be a legitimate triplet from the specific triangle assignment concerned. Adding these up, their total is

$$A_{26}+A_{36}+A_{66}+A_{86}+A_{96} = [(2,3,6)+(3,8,6)+(8,9,6)+(9,2,6)]/3$$

which is exactly the area (quadrature factor) A_6 as specified on page 79. This demonstrates the assertion in this particular example. The general case is not difficult to handle.

The ordered ensemble λ_i^* will start with reasonably large values λ_1^*, λ_2^*, --- but these will decrease rapidly. On the other hand, the ordered ensemble of proper values, μ_i, will generally not be nearly as large as those of λ_i^* to start, but will decrease rather slowly (the nature of the decrease will depend on the ratio of the number of observations to the order of the matrix concerned). As a consequence, the ordered proper values of the total matrix K_{ij}, λ_i, will approximate the values of λ_i^* for small i, but will be dominated by the proper values μ_i for larger values of i. The "change-over" point will be in the neighborhood of the "knee" of the curve of log λ_i vs i as illustrated in Figures IV-8 and IV-9 but cannot be exactly located there. The effect of the "smoothing" property of the product-integral technique is to drastically reduce the effect of the proper values of the matrix of departure from the true covariances, k_{ij} (i.e., all of the proper values μ_i are reduced in size by smoothing) while the effect on the true values λ_i^* is much smaller. The net effect is then to decrease the size of the larger computed total proper values, λ_i, when the product-integral technique is used.

Now that the smoothing of the covariance kernel in the product-integral technique is established, we proceed to consider the effect that this would have on the proper values and proper functions. Since we are dealing with an empirically determined function, an important consideration is the fact that every covariance value K_{ij} is affected by sampling variations and the values K_{ii} (i.e., when i=j) contain the residual variances in addition to the variance of the true values. The next section is devoted to the method for accounting for residual variances, but they have not been accounted for at this point. Thus, the matrix of observed variances and covariances may be written as the sum of two matrices $K_{ij}=K^*_{ij}+k_{ij}$ where K^*_{ij} are the variances and covariances of the "true values" and k_{ij} are the departures from the true values (which on the diagonal i=j may be quite large). Now the proper values of the matrix K_{ij}, say λ_i, are associated with the proper values of the matrix K^*_{ij}, say λ^*_i, and the proper values of the matrix k_{ij}, say μ_i, but we cannot write $\lambda_i= \lambda^*_i+k_i$ where λ_i, λ^*_i, μ_i are all three ordered. We can only write $\lambda_i = \lambda^*_a + \mu_b$ where λ_i is ordered, $\lambda_1>\lambda_2>\lambda_3---\lambda_n$ and the subscripts a and b fall in some order to suit the situation. It will be generally true that the effect of the residual variances and sampling variation on the proper functions for the trapezoid method as compared with the product-integral technique is of a somewhat different character. These are dependent in a large measure on the nature of the sampling variation displayed by k_{ij} at the off-diagonal points, $i\neq j$. In the product-integral technique these values are strongly smoothed while in the trapezoid method they are not. As a consequence, one would expect that there would be (after a certain undetermined index value) a larger difference (smaller correlation) between proper functions computed by the different quadrature methods for the same triangle assignments than for different triangle assignments for the same quadrature method. This is, of course, what is being illustrated in Figures IV-10 and IV-11.

The conclusion to be drawn from the above comparison of two quadrature methods for two apparently equally valid triangle assignments is that there are certainly no more than 7 (or possibly 6) proper functions that are adequately self-consistent and that if one wishes to take a more conservative attitude, this can be reduced to only 4.

3. Factor Analysis Methods

In the preceding section it was shown that both the assignment of the triangle coverage of the region concerned and the quadrature method used to evaluate the integral had a strong effect on the self-consistency of the proper functions computed. It was also pointed out there that at least one reason for this was due to the effect of sampling variation and the fact that the residual variances had not been removed. Of these two, one may at least approximate the residual variances and remove their effect by using Factor Analysis methods. This approach is considered in this section.

In order to obtain an estimate of the residual variances contained in the St. Louis SO_2 data, resort was made to the abbreviated method due to Jöreskog (1962). The method of Jöreskog was modified in that, in order to keep the advantages of the integral equation formulation of the problem, it was applied after the equations were formulated in a symmetrical matrix form. (That is, the covariance matrix was modified by the proper quadrature factors for the trapezoid method; the straight principal component formulation based on "equal quadrature factors for all points" was not used.)

The comparison of proper functions for the two methods is shown in Figure IV-12. It is to be noted in this figure that the covariance of proper functions for the two triangle subdivisions is reasonably high up through the 8'th proper function, after which it takes an abrupt dip. On the basis of the argument that once this dip occurs, the proper functions are essentially irrelevant due to the mathematical

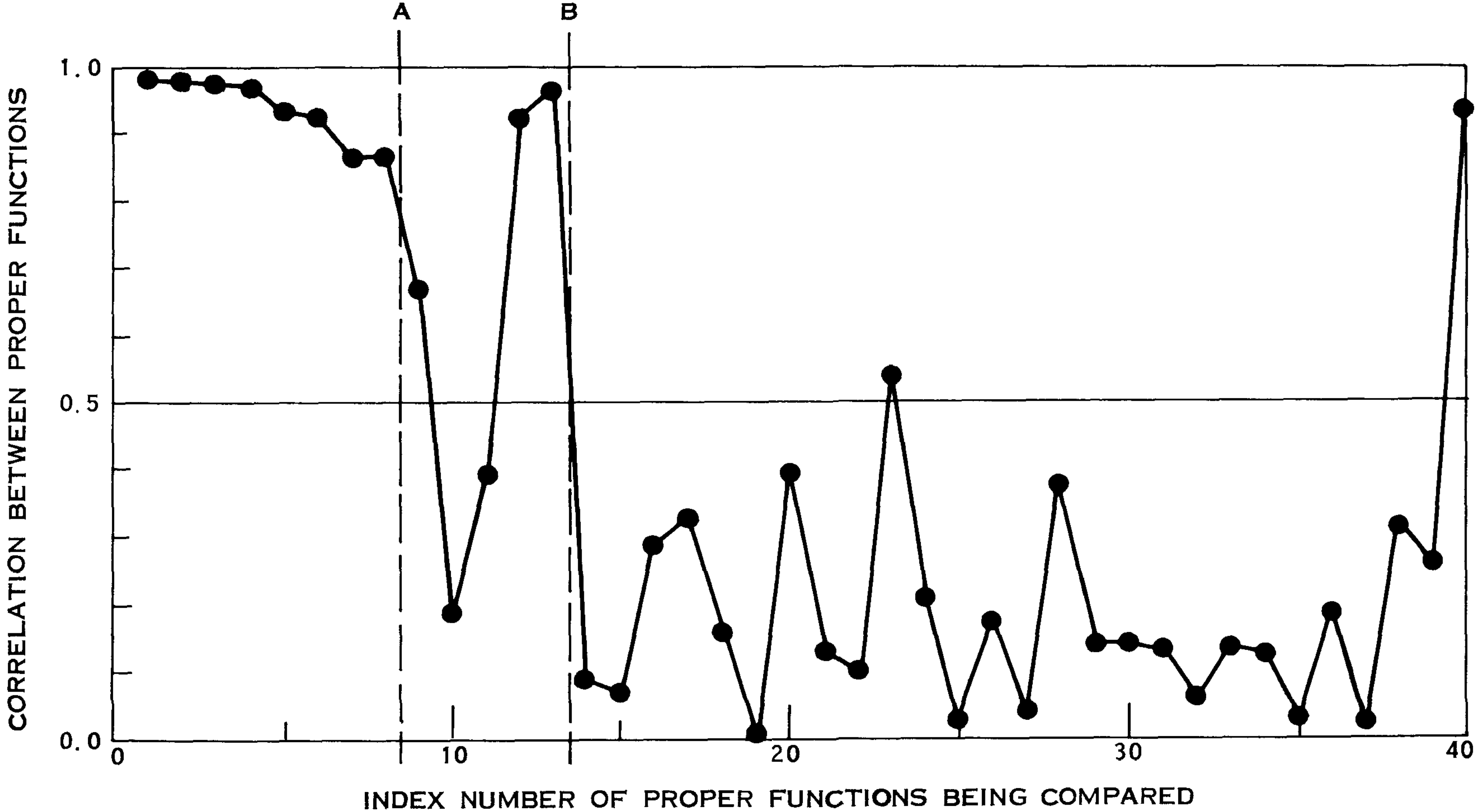


FIGURE IV–12

CORRELATION OF PROPER FUNCTIONS FOR TWO DIFFERENT BUT ACCEPTABLE ASSIGNMENTS OF QUADRATURE PARAMETERS. DUE TO SIGN AMBIGUITY ONLY NUMERICAL VALUES PLOTTED. LINE A INDICATES FIRST SIGNIFICANT DROP IN CORRELATION, LINE B INDICATES CUTOFF BASED ON 5% SIGNIFICANCE LEVEL USING JORESKOG'S METHOD.

(not statistical) ambiguities of the quadrature formula, we can assert that there are not more than 8 significant proper functions. This is indicated by the line A in Figure IV-12.

On the basis of the analysis of Jöreskog's method using the 5% level, there are at least 13 significantly different proper values. This is indicated by the line B in Figure IV-12. It is to be noted that the drop in the correlation of the values of the proper functions between 8 and 13 is apparently transient in that the correlation between functions 12 and 13 is as high as that for those of index less than 8. It would appear that the statistical test for significantly different proper values is far less restrictive than the test based on the ambiguity of the mathematical problem. After the 13'th proper function has been passed, the corresponding proper functions are very poorly correlated, uniformly until the last one is reached.

The correlation of proper functions shown in Figure IV-12 corresponds to that of Figure IV-10 (the points indicated by the dots) except for the fact that in the case of Figure IV-12 the proper functions correspond to the weighted covariance matrix modified by multiplying ahead and behind by the square root of the diagonal of the inverse of the weighted covariance matrix as described on page 66. As shown there (page 66 and following) this has the effect of making the resulting "residual variances" (that appear on the diagonal after modification) all equal. It is well known that if a positive quantity is added to the diagonal of a symmetric matrix, all of the proper values are increased by this amount while the proper functions remain unchanged (Anderson, 1958). Figure IV-12 then implies that after the Jöreskog modification (as described on page 66) the effect of the two different triangle assignments is to effectively make all corresponding (same index) proper functions after the 13'th essentially uncorrelated, while those with index 13 or less are highly correlated (the three exceptions are discussed later). This, of course, is primarily due to sampling variations alone since the effect of the residual variances has been made uniform after the matrix modification. One is thence led to conclude that the early "cut off" of the correlation between corresponding proper functions shown in Figure IV-10 was due to the fact that the point to point variation of the residual variances was a dominant factor.

To summarize the situation in different terms, the computation of the proper functions and proper values without removing the residual variances (more exactly, without modifying the covariance matrix weighted by the quadrature factors to give a uniform equivalent of the residual variances) results in proper functions that become inconsistent (low correlation of corresponding proper functions for different weight factors of equal validity) at a very low index and for which the consistency check (correlation) behaves in a very irregular way (sometimes high, sometimes low) for the higher index numbers. On the other hand, if the Jöreskog method is used (wherein the weighted covariance matrix is modified to obtain a uniform equivalent of the residual variances) then the consistency check (correlation of proper functions for different weight choices) clearly divides the proper functions into two distinct groups; those with index lower than a certain value (here, 13) that are rather uniformly self-consistent (highly correlated for different weight choices), and those with higher index value (here, 14 and above) that are uniformly inconsistent (low correlation for different weight factor choices). This is precisely the result that we had hoped to achieve.

The low values of correlation between proper functions at index values 9, 10, 11 of Figure IV-12 are due to the fact that the matrix method does not give self-consistent results for the solution of the integral equation type problem when different quadrature factors are used. In other words, the modification of the original matrix by the quadrature factors has introduced considerations that are not accounted for by the sole question of statistical significance.

It is to be noted that the Jöreskog test is based on proper values alone while that of the correlation test is based on the proper functions alone. On the other hand, applications are all essentially based on the proper functions themselves. As a consequence, it is necessary to consider further the question of the application of the proper functions 9-13.

To see more clearly what is taking place in the consistency check for different quadrature factors (weights) the elements near the principal diagonal of the matrix

$$C_{ij} = \sum_k \theta^*_i(x_k)\,\theta^\#_j(x_k)$$

are tabulated in Table IV-4. In this table the central column headed "same" contains the values C_{ii} on the principal diagonal while the column on the left headed "order" is the value of i. The tabulated values are thence the correlation of the proper functions of the same index but with different but equally valid quadrature factors (weights). The diagonals of the matrix C_{ij} that border the principal diagonal appear in the adjoining columns of Table IV-4. The values of $C_{i,i+k}$ are listed under the columns headed by the values k=-3,-2,-1, +1,+2,+3. The absolute values of the numbers in the column "same" are those shown in Figure IV-12.

Table IV-4. Correlation Between Proper Functions for Different Area Assignments. The central column indicates proper functions of same order. Adjacent columns are the correlations between proper functions of different orders.

Order	-3	-2	-1	Same	+1	+2	+3
1	---	---	---	.982	-.005	+.049	-.013
2	---	---	.002	.974	.010	-.061	.015
3	---	-.050	.007	.974	-.007	+.031	-.009
4	.013	.062	.006	.967	.133	.014	-.072
5	.003	-.027	-.134	.935	-.016	.024	.229
6	.001	-.025	.082	.928	.087	-.242	.139
7	.045	.001	-.176	.858	-.212	.254	.223
8	-.189	.226	.308	.862	-.091	-.128	-.041
9	.070	.050	-.017	-.671	.674*	-.092	-.191
10	.177	-.103	-.297	-.193	.880*	.028	-.018
11	.200	-.585*	-.575*	-.391	.184	.024	-.026
12	.044	-.231	-.043	-.929	.060	.009	.030
13	-.054	.039	.074	.965	.057	.025	.025
14	-.011	.015	-.002	-.098	.099	-.207	.534
15	-.038	.095	-.824	.072	-.011	-.103	.068
16	-.021	-.002	-.040	-.292	-.199	.398	.047
17	.055	-.529	.007	-.338	.057	-.037	.062
18	-.011	-.118	-.037	.163	-.027	-.162	.190
19	.114	-.061	-.445	-.004	.341	-.018	.115
20	.075	.198	-.367	.397	.445	.024	-.017

* See text for a discussion of these values.

In Table IV-4 the large off-diagonal correlations for proper functions 9, 10 and 11 are marked with an asterisk, *. It is indicated that proper function 9 is highly correlated with itself and 10, 10 with 11, and 11 with 10 and 9. In other words, as the weights (quadrature factors) are changed the proper functions 9, 10 and 11 shift about among themselves. In more detail, if $\theta_i^*(x_k)=\theta_i^*$, $\theta_i^\#(x_k)=\theta_i^\#$ are the i'th proper functions for two different but equally valid quadrature factor assignments, we have the approximate relations

$$\theta_9^\# \approx -0.671\theta_9^* + 0.674\theta_{10}^*$$

$$\theta_{10}^\# \approx \qquad\qquad\qquad\qquad + 0.880\theta_{11}^*$$

$$\theta_{11}^\# \approx -0.585\theta_9^* - 0.575\theta_{10}^*$$

(The relations would be exact if all 40 proper functions were listed on the right with appropriate coefficients from the full table.) Thus θ_{11}^* shifts to $\theta_{10}^\#$ while θ_9^* and θ_{10}^* go into $\theta_9^\#$ and $\theta_{11}^\#$ (via a rotation of about 135°). (All of this is, of course, dependent on the relations between the two quadrature factor selections and would be invalid for any other pair of such quadrature factors.) Further information on this phenomena is provided by the proper values of the modified weighted covariance matrix which are illustrated in Figure IV-13. It is seen there that proper values 9 and 10 are very nearly equal.

When two (or more) proper values of a matrix (symmetric, positive definite) are equal, the proper functions corresponding to these are not both uniquely defined (Anderson, 1958). Apparently this pair of proper values are sufficiently close to each other that the change in the quadrature factors used was sufficient to show up this "near indeterminancy".

Table IV-4 for order numbers 14-20 shows the small correlations for C_{ii} of Figure IV-12 in the column headed "same", generally small correlations in the off-diagonal positions, but a few scattered larger values but no apparent pattern.

The residual variances depend strongly on the point at which the number of significant proper values/functions is terminated. These are shown in Table IV-5 for a termination of 8 or 13 such. Only the values for one subdivision of the area into non-overlapping triangles is shown since to the number of digits shown in this table, the results of the two subdivisions into triangles were identical. It is to be noted that, since SO_2 concentrations are approximately log-normally distributed, the variances are those of the logarithm of the SO_2 concentration. The station-to-station variation of the fraction of the residual variance compared with the total residual variance is to be noted. These are particularly large at stations 2, 9, 10, and 30. It is presumed that this is due to station instrumentation or instrument exposure at these locations since these locations are not obviously related to each other (i.e., other stations with relatively small residual variances lie between each pair (see Figures IV-5, 6, 7).

The residual variance computed for station 9 clearly indicates that all 13 of the proper functions are significant in the case at hand. If only the first 8 proper functions are used, the computed residual variance is larger than the total variance, which is quite impossible.

It was pointed out above that the residual variance at locations 2, 9, 10, 30 appear to be quite large. A remark on the significance of the ratio of residual variance to

Table IV-5. Total Variances, Residual Variance and the Fraction of the Total Variance for St. Louis SO_2 Data at 40 Stations. The abbreviation P.V./F. stands for proper values/functions. The number preceding this abbreviation indicates the number of P.V./F.'s used in computing the residual variances shown. See text, p. 118.

Station	Total Variance	8 P.V./F. Residual Variance	8 P.V./F. Fraction	13 P.V./F. Residual Variance	13 P.V./F. Fraction
1	.0497	.0049	.10	.0026	.05
2	.0829	.0351	.42	.0183	.22
3	.0374	.0072	.19	.0037	.10
4	.0583	.0121	.21	.0063	.11
5	.0269	.0029	.11	.0015	.06
6	.0982	.0151	.15	.0079	.08
7	.0738	.0086	.12	.0045	.06
8	.0784	.0048	.06	.0025	.03
9	.0945	.1050	1.11	.0548	.58
10	.0376	.0208	.55	.0109	.29
11	.0295	.0026	.09	.0014	.05
12	.0758	.0028	.04	.0014	.02
13	.0352	.0049	.14	.0026	.07
14	.1118	.0229	.20	.0120	.11
15	.0976	.0114	.12	.0059	.06
16	.1938	.0184	.09	.0096	.05
17	.0269	.0060	.22	.0031	.12
18	.0532	.0051	.10	.0027	.05
19	.1155	.0341	.30	.0178	.15
20	.0567	.0090	.16	.0047	.08
21	.0778	.0040	.05	.0021	.03
22	.0684	.0074	.11	.0038	.06
23	.0854	.0087	.10	.0045	.05
24	.1468	.0159	.11	.0083	.06
25	.1527	.0139	.09	.0073	.05
26	.1183	.0071	.06	.0037	.03
27	.1352	.0070	.05	.0036	.03
28	.1765	.0378	.21	.0197	.11
29	.1195	.0047	.04	.0024	.02
30	.1067	.0752	.70	.0392	.36
31	.1100	.0111	.10	.0058	.05
32	.0743	.0129	.17	.0067	.09
33	.1299	.0186	.14	.0097	.07
34	.0976	.0126	.13	.0066	.07
35	.0861	.0222	.26	.0116	.13
36	.1051	.0141	.13	.0074	.07
37	.1114	.0059	.05	.0031	.03
38	.0570	.0126	.22	.0066	.12
39	.1016	.0126	.12	.0066	.06
40	.0428	.0125	.29	.0065	.15

total variance may be appropriate at this point. The ratios in the column headed fraction, F, are the values

$$F = \sigma_r^2/(\sigma^2 + \sigma_r^2)$$

where σ_r^2 is the residual variance and σ^2 is the "true" variance. The ratio of the residual to "true" variance is given by

$$\sigma_r^2/\sigma^2 = F/(1-F).$$

The ratio of the standard deviation of the residuals to the "true" standard deviation is the square root of the above. Some values are tabulated in Table IV-6.

Table IV-6. Ratio of Standard Deviation of Residuals (σ_r) to the Standard Deviation of "True" Values (σ) as a Function of the Ratio of Residual Variance to Total Variance (F)

F	σ_r/σ	F	σ_r/σ
0.00	0.00	0.30	0.66
0.05	0.23	0.40	0.82
0.10	0.33	0.50	1.00
0.15	0.42	0.60	1.22
0.20	0.50	1.00	∞

It is to be noted that in the case of the four locations noted above all have a value of F in excess of 0.20 which means that the standard deviation of the residuals is more than 50% of the standard deviation of "true" values. Table IV-5 lists 13 locations with F values of 0.10 or more for which the standard deviation of residuals is more than 33% of the standard deviation of "true" values. These seem to be rather large.

The significance of the residuals should also be noted here. In the type of analysis that has been made, the term residual includes not only the errors of observation, but also the ability of the station network to resolve small scale effects. Thus, the size of the residuals is dependent on the density of the pollutant concentration measurement points.

It was pointed out earlier that the proper functions were not comparable with each other for various triangle subdivisions or quadrature techniques after a certain index had been reached and that this index seemed to be associated with the index at which a "knee" appeared in the plot of log-proper value against index number. The proper values of the matrix after being modified by multiplying ahead and behind by the square root of the diagonal of the inverse matrix, an essential feature of the Jöreskog technique, are shown in Figure IV-13. The numerical values of the proper values have been radically changed, but the shape of the curve of the logarithm of the proper value against its index remains about the same with the exception that the "knee" has been removed and is replaced by a gradual change of slope between index numbers 5 and 10.

In view of the preceding analysis, we would then conclude that the criterion of Craddock and Flintoff (1970) and Farmer (1971) is not really a measure of the significant number of proper values/functions but is a phenomenon associated with the fact that the residual variances have not been adequately treated by a principal component analysis. The Factor Analysis procedures used here give different results that seem to be self consistent and which in addition provide a quantitative estimate of the residual variances that were hitherto ignored.

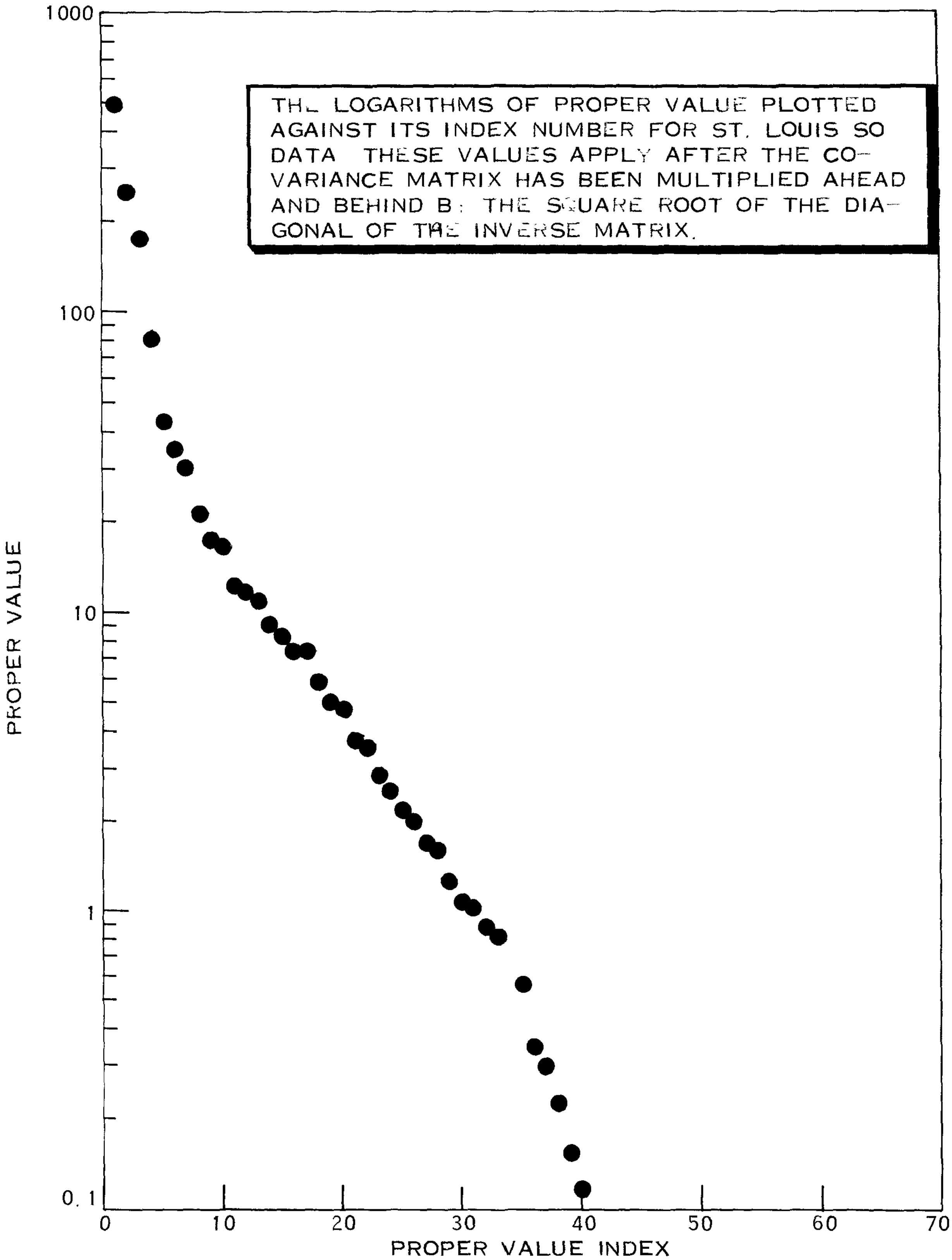

THE LOGARITHMS OF PROPER VALUE PLOTTED AGAINST ITS INDEX NUMBER FOR ST. LOUIS SO DATA THESE VALUES APPLY AFTER THE CO-VARIANCE MATRIX HAS BEEN MULTIPLIED AHEAD AND BEHIND B: THE SQUARE ROOT OF THE DIAGONAL OF THE INVERSE MATRIX.
PROPER VALUE
1000
100
10
1
0.1
0
10
20
30
40
50
60
70
PROPER VALUE INDEX

FIGURE IV-13

Note that in Figures IV-10, IV-11 and IV-12 the correlation of the 40'th proper function for different quadrature methods (Figure IV-11) and for different quadrature factors (Figures IV-10 and IV-12) is extremely large and is an isolated point in Figure IV-12. This is a mathematical phenomena and has nothing to do with the fields involved. As a matter of practical experience, the larger the index, the more oscillatory the proper functions. This is strictly true of what has been called an "oscillatory" matrix (Gantmacher, 1960b). We have not found a proof for any other type of matrices and one can in fact construct symmetric positive definite matrices for which the opposite is true. We strongly suspect that "oscillatory" matrices *do not* exhaust the class of matrices with this property, but characterize a particular class of matrices for which this property could be proved. With the fact that experience indicates a highly oscillatory behavior for the last proper function, we also suspect that there is a tendency for these oscillations to either be in or out of phase for matrices that are similar to each other, as is the case for the matrices concerned here. This would account for the high correlation observed for proper functions No. 40.

D. Summary and Conclusions

In this chapter, the subject of the analysis of covariance matrices has been treated in some detail, since it is of great importance in the analysis of data fields of all kinds and in particular to the fields of pollution concentration data, especially in connection with the problem of optimum observation station locations to which it is applied in the computer program that is discussed in detail in the final chapter. Section A was devoted to a general look at the problem, Section B to a detailed discussion of how to determine the residual variances and the number of significant proper values and proper functions required to describe the statistical properties of the pollutant concentration field using Factor Analysis methods but modified by the fact that one is dealing with a two dimensional field and should use a basic integral equation formulation to account for variable data density. The techniques developed in Section B were applied to St. Louis SO_2 concentration data in Section C.

As a result of the analysis in Section C, it was found that the Factor Analysis technique as modified by the integral equation formulation gave (1) a quantitative evaluation of the number of significant proper values and proper functions required to describe the pollution field, (2) a quantitative evaluation of the residual variances in such detail that unique values were assigned each observation location separately (much more than a general estimate), (3) that the method was consistent in spite of the fact that there are mathematical ambiguities in the quadrature process (evaluation of the integrals involved), and (4) that the self-consistency of the solution from the mathematical point of view corresponded very closely to the tests for significance from the statistical point of view. It was further concluded that the Factor Analysis approach is a <u>prime</u> <u>requisite</u> for the success of the analysis since when this approach is disregarded the results tests for mathematical self-consistency and for statistical significance are not completely compatible with each other; and it was shown that this results from ignoring the importance of the residual variances.

CHAPTER V

PROGRAM BAST AND ITS SUBROUTINES

The program for an optimal location of pollution observation points (BAST) and its twelve associated subroutines are described in some detail in the following sections. The interrelations between these subroutines and the main program is shown in the following table. The order in which they are discussed follows

TABLE 1

THE RELATIONS BETWEEN PROGRAM BAST AND ITS SEVERAL SUBROUTINES

SUBROUTINES CALLED	PROGRAM	CALLING SUBROUTINE						
	BAST (A)	CIRCUM[2]	FMFP[3]	FUNB[4]	FUNCT[5]	ADDPT[6]	TRIFIX[7]	PORDR[8]
CIRCUM (B)	√	=						
FMFP (C)	√	√	=					
FUNB (D)	√		√	=				
FUNCT (E)	√		√	√	=			
CORFUN (F)					√			
INT2D (G)					√			
MATINV (H)					√			
ADDPT (I)	√	√				=		
TRIFIX (J)						√	=	
PORDR (K)						√		=
AR2 (L)						√	√	√
TRITST (M)						√		

that indicated by the letters in parentheses in the subroutines called column, BAST itself being considered first.

As shown in the table, the subroutines are divided into two categories: those involved in computation of the station location [(B) through (H)], and those involved with bookkeeping required to preserve, at all times, a satisfactory subdivision of the area concerned into completely covering but non-overlapping triangles [(I) through (M)].

A. PROGRAM BAST

This program is set up to accept NSTN coordinates XS(I), YS(I), I=1, NSTN, of already located observation stations of which NBDY of these form the convex boundary of the area covered by these station locations (line 38). The index numbers of the boundary stations are entered as IBDY(I), I=1, NBDY, and should be in counterclockwise order around the boundary but with any initial starting point (line 41). The boundary of the area covered by the initial stations is a convex polygon with the boundary stations at the vertices. The area is convex in the sense that it must not have any re-entrant corners. Put in a different way, if one traverses the boundary in the counterclockwise direction, then at each vertex, the line segment that forms the boundary is rotated in a counterclockwise direction. (The mathematical statement is that the area is convex if the line joining any pair of points in the area lies entirely within the area.)

The area covered by the existing station location is subdivided into non-overlapping triangles with stations at the vertices (line 52). The boundary segments are sides of some of these triangles. There are NTR such triangles. The index numbers of their vertices are ITR(I,1), ITR(I,2), ITR(I,3), I=1, NTR, and these must be in counterclockwise order about each triangle.

The entire region is interior to a circle with center at (XC,YC) and of radius R (line 56). Some of the stations listed previously may lie on this circle. If this is the case, their azimuth must be specified. The number of such is NCIR and the point azimuth AZ(I), I=1, NCIR is measured counterclockwise (degrees) from East.

Additional information to be specified (line 81) are: the number of additional stations to be located (NADD), the percent reduction required between successive maximum errors of estimate (PMIN), the expected absolute error of the minimization subroutine (EPS), the maximum number of iterations allowed for this subroutine (LIMIT), and a control for print-out of details (IPRNT).

Since the program calls for computation of correlation coefficients via their proper value/function representation at a grid of points covering a square 80 km on a side, the coordinates of the grid points (lines 87,88), the proper values (line 93) proper functions (line 96), and standard deviations for weight parameters (line 101) are required inputs. The program, as it stands, is set up for 15 proper values/functions on a 9x9 grid with 10 km spacing between points. The weight parameter (line 101) corresponds to the real statistical situation where expected standard deviations of pollutant concentrations are used, but the weight parameter may be arbitrary. The program assumes that it may be differentiated using finite (1 km) differences with sufficient accuracy.

If less than two existing stations are listed on the circle that defines the region being considered, then two such points are located on the circle (lines 114-135).

The criterion for station location is to the effect that a "best" location is at a point where the error of estimate using a linear least squares regression on sample values at existing stations would be a maximum. Since such errors of estimate are bounded between zero at an existing station and the variance of pollution concentration, it is reasonable to look for such maxima at locations distant from existing stations. Consequently, candidate locations are taken to be (a) at the center of gravity of the station triangles or (b) midway on the circle between stations located on the circular boundary (lines 135-173). Of the errors of estimate found at these locations, the largest three are selected and the neighborhood of each is searched to find the value of the local maximum in its neighborhood. These local maxima are then compared and the new station location assigned at that point which has the largest local maximum (lines 174-252). A bit of manipulation is required (lines 253-275) to insure that new points on the circular boundary do not wander outside the circle.

The total number of stations located (NSTN) is then checked against the total number required (NSTOP) (line 287). If this number has not been reached the process is repeated by returning to line 135. If the required number has been reached the remainder of the program (lines 290 through 322) prints out the results and terminates. The final lines 323 through 345 are devoted to various diagnostics.

Note: The subroutine FMFP is a standard subroutine that locates the <u>minimum</u> of a function of several variables. The procedure described above involves locating a <u>maximum</u>. To "trick" FMFP into thinking a maximum is a minimum, the mean square error of estimate is given (internally) a negative sign. Whenever the term <u>minimum</u> appears in subsequent descriptions it is to be understood in this sense.

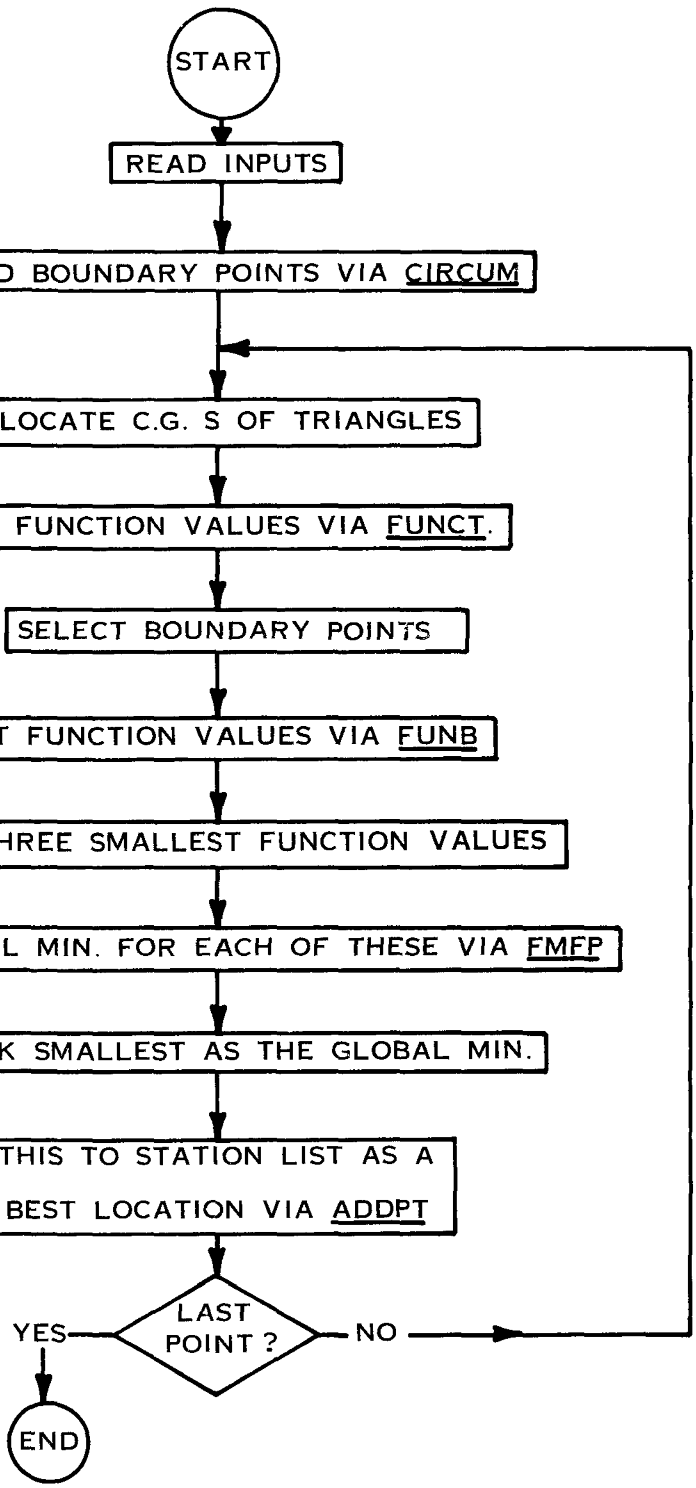


FIGURE V-1

ABBREVIATED FLOW CHART OF MAIN PROGRAM BAST

```
*DECK MAIN
      PROGRAM BAST( INPUT, OUTPUT )
C     DETERMINES LOCATION OF POINTS OF LARGEST ERROR OF ESTIMATE FROM
C     EXISTING NET AND LOCATES NEW STATION THERE.  PROCEDURE ITERATED
C     UNTIL EITHER TOTAL NO. OF STATIONS REACHES PRESCRIBED SIZE OR
C     TILL THE PERCENT REDUCTION OF THE MAX EOFE IS BELOW PRESCRIBED
C     AMOUNT
      COMMON / BLK1 / NSTN, NBDY, NTR, XS(30), YS(30), ITR(50,3)
     , , XC, YC, R, IBDY(20)
      COMMON / BLK2 / XD(9), YD(9), PH(15,9,9), ALAM(15), W(9,9)
      COMMON / BLK3 / NCIR, AZ(20)
      COMMON / BLK4 / EST, EPS, LIMIT, IER, F
      COMMON / LINKS / ERR1, X0(60), DIFF(60)
      COMMON / Q / IC, IQ
      DIMENSION        X(2), XT(60),
     , YT(50), SAMP(60), ITRY(3), JTRY(3), STAMP(3), XMAX(3), YMAX(3),
     , FMAX(3), FM2(3), XA(3), YA(3),       XTMP(3), YTMP(3)
      DIMENSION G(2)
      DIMENSION TITLE(32)
      EXTERNAL FUNCT, FUNB
      DATA ((TITLE(I),I=1,32)=
     1    10HPARAMETER ,10HA1, FIRST ,10HCOEFFICIEN,10HT           ,
     2    10HPARAMETER ,10HA2, SECOND,10H COEFFICIE,10HNT          ,
     3    10HPARAMETER ,10HA3, ORIGIN,10H IN FIRST ,10HTERM        ,
     4    10HPARAMETER ,10HA4, ORIGIN,10H IN SECOND,10H TERM       ,
     5    10HPARAMETER ,10HA5, FIRST ,10HSCALE FACT,10HOR          ,
     6    10HPARAMETER ,10HA6, SECOND,10H SCALE FAC,10HTOR         ,
     7    10HPARAMETER ,10HA7, THIRD ,10HSCALE FACT,10HOR          ,
     8    10HPARAMETER ,10HA8, ORIENT,10HATION ANGL,10HE           )
      DATA IC, IQ / 1, 0 /
C     DATA INPUT
C     NSTN=NUMBER OF LOCATIONS
C     XS(I), YS(I), I=1,NSTN, STATION COORDINATES, KM FROM ARB. ORIGIN
C     NBDY = NUMBER OF BOUNDARY POINTS
C     THE BOUNDARY MUST BE CONVEX, BOUNDARY POINTS ARE NUMBERED FROM
C     NSTN - NBDY + 1 TO NSTN AND ARE LISTED IN COUNTER CLOCKWISE ORDER
C     ALL OTHER STATIONS MUST LIE INSIDE  ( NOT ON ) THE BOUNDARY BUT
C     THE ARRANGEMENT OF INDEX NUMBERS IS IRRELEVANT
      READ 1052, NSTN, ( XS(I), YS(I), I = 1, NSTN )
 1052 FORMAT (I5/(2F10.0))
C     LIST OF BOUNDARY POINT NUMBERS IN CC ORDER AROUND THE BOUNDARY.
      READ 1053, NBDY, ( IBDY(I), I = 1, NBDY )
 1053 FORMAT (16I5)
C     SUBDIVIDE REGION INTO TRIANGLES
C     NTR = NUMBER OF TRIANGLES
C     ITR(I,J), I = 1, NTR, J = 1, 3, INDEX NUMBERS OF THE VERTICES
C     1, 2, 3, OF TRIANGLE NO. I.  THE NOS. 1, 2, 3, ARE TO BE IN CC
C     ORDER AROUND EACH TRIANGLE.  TRIANGLE SIDES THAT ARE ALSO SIDES
C     OF THE BOUNDARY MUST HAVE 1,2; 2, 3; OR 3, 1 IN THE ORDER IN
C     WHICH THESE POINTS APPEAR ON THE BOUNDARY.  THE ENTIRE INTERIOR
C     OF THE AREA MUST BE COVERED.  TRIANGLES MUST NOT OVERLAP.
C     NUMBERING OF THE TRIANGLES IS NOT SIGNIFICANT.
      READ 1054, NTR, ( ITR(I,1), ITR(I,2), ITR(I,3), I = 1, NTR )
 1054 FORMAT (I5/(3I5))
C     THE OUTER CIRCLE, CENTER AT XC, YC, AND RADIUS R
```

```
C     NCIR = NUMBEF OF POINTS ON THE CIRCLE
      READ 1058, XC, YC, R, NCIR
1058  FORMAT (3F10.0,I5)
C     AZIMUTH OF POINTS ON THE CIRCLE
C     AZ(I), I = 1, NCIR, AZIMUTH IN DEGREES CC FROM EAST
C     THESE MUST BE BOUNDARY POINTS APPEARING IN THE LISTS XS(I), YS(I).
C     NONE OF THE BOUNDARY POINTS MAY LIE OUTSIDE OF THE CIRCLE.
C     SOME OR ALL MAY LIE INSIDE.  IF ALL LIE INSIDE THEN NCIR = 0.
C     THEY MUST APPEAR IN ORDER OF INCREASING AZIMUTH, 0 TO 360 DEGREES.
      IF ( NCIR .LE. 0 ) GO TO 15
      READ 1060, ( AZ(I), I = 1, NCIR )
1060  FORMAT ( 9F8.0)
      DO 1  I = 1, NCIR
1     AZ(I) = AZ(I) * .017453292
C     MISCELLANEOUS OTHER INPUT PARAMETERS
C     NADD = NUMBER OF POINTS TO BE ADDED.
C     IF NCIR = 0, 1, THEN 2 OR 1 POINTS WILL BE ADDED ON THE CIRCLE
C     REGARDLESS OF NADD.  IF THIS IS NOT THE STOPPING CRITERION, NADD=0
C     PMIN = PERCENT REDUCTION BETWEEN SUCCESSIVE MAXIMUM ERRORS OF
C     ESTIMATE.  NO POINTS WILL BE ADDED AFTER PERCENT REDUCTION GOES
C     BELOW THIS VALUE.
C     EPS = EXPECTED ABSOLUTE ERROR FOR SUBRPOUTINE FMFP
C     LIMIT = MAX. NO. OF ITERATIONS FOR SUBROUTINE FMFP
C     SUGGESTED VALUES FOR EPS = .01, AND LIMIT = 5
C     IPRNT = REQUEST FOR PRINTOUT OF DETAILS OF THE INPUT OTHER
C     THAN COORDINATES OF INITIAL STATIONS
15    READ 1062, NADD, PMIN, EPS, LIMIT, IPRNT
1062  FORMAT (I5,F10.0,E10.1,2I5)
      TPI = 6.2831853
      NSTOP = NSTN + NADD
C     COORDINATES FOR THE CORRELATION COEFFICIENT
C     PARAMETERS XD(I), YD(I), I = 1, 9
      READ 1056, ( XD(I), I = 1, 9 )
      READ 1056, ( YD(I), I = 1, 9 )
C
C     CORRELATION COEFFICIENT DATA INPUT IN TERMS OF
C     PROPER VALUES AND PROPER FUNCTIONS
C
      READ 1055, ( ALAM(I), I = 1, 15 )
1055  FORMAT (8F10.0)
      DO 2 I = 1, 15
2     READ 1057, ( ( PH(I,J,K), K = 1, 9), J = 1, 9 )
1057  FORMAT ( 16X,4E16.8/5E16.8/(4E16.8/5E16.8))
1056  FORMAT (9F8.0)
C     WEIGHT FUNCTION TABLE FOR VALUES AT POINTS XD(I), YD(J)
      DO 3  J = 1, 9
      READ 1056, ( W(I,J), I = 1, 9 )
3     CONTINUE
C     OUTPUT OF THE INPUT STATIONS
2000  FORMAT (1H1)
2002  FORMAT (1H0)
      PRINT 2000
      PRINT 2010
2010  FORMAT (5X,*A. COORDINATES OF INITIAL STATIONS*/5X,6HSERIAL,2X,
     111HCOORDINATES/5X,6HNUMBER,5X,1HX,5X,1HY)
```

```
      PRINT 2012, ( I, XS(I), YS(I), I = 1, NSTN )
 2012 FORMAT (I10,F7.1,F6.1)
C     HEADING FOR RESULTS
      FREF = 1.
      IF ( NCIR .GT. 1 ) GO TO 8
C     ADDS ONE OR TWO POINTS ON THE CIRCLE
      IQ = 1 $ IC = 2
      CALL CIRCUM( XC, YC, R, NCIR )
      PRINT 2018
 2018 FORMAT (5X,*B. COORDINATES OF CIRCLE POINTS ADDED*)
C     PRINTS COORDINATES AND SERIAL NUMBERS OF POINTS ADDED ON CIRCLE
      DO 6 I = 1, NCIR
      I1 = NSTN - NCIR + I
      PRINT 2012, I1, XS(I1), YS(I1)
 6    CONTINUE
      PRINT 2014
 2014 FORMAT (5X,*C. COORDINATES OF ADDED POINTS AND SUPPLEMENTARY DATA*
     ))
      PRINT 2016
 2016 FORMAT (5X,6HSERIAL,2X,11HCOORDINATES,5X,6HE OF E,2X,7HPERCENT,5X,
     ,21HALTERNATIVES REJECTED/5X,6HNUMBER,5X,1HX,5X,1HY,6X,6HBEFORE,2X,
     ,7HREDUCT.,5X,1HX,5X,1HY,4X,6HE OF E,7X,1HX,5X,1HY,4X,6HE OF E)
      ISPACE = 0
C     SAMPLES POINTS FOR ERRORS OF ESTIMATE
C     FIRST GETS POINTS AT CG OF TRIANGLES
      EST = -1.
 8    DO 12 I = 1, NTR
      SUM1 = SUM2 = 0.
      DO 10 J = 1, 3
      SUM1 = SUM1 + XS(ITR(I,J))
 10   SUM2 = SUM2 + YS(ITR(I,J))
      X(1) = XT(I) = SUM1 / 3.
      X(2) = YT(I) = SUM2 / 3.
      IQ = 2 $ IC = 1
      IF ( I .EQ. 1 ) IQ = 1
      EST = .9 * F - .1
      CALL FUNCT( 2, X, F, G )
 12   SAMP(I) = F
C     SELECTS POINTS ON THE CIRCLE
      DO 14 I = 1, NCIR
      I1 = I + 1
      IF ( I .EQ. NCIR ) I1 = 1
      ANG = ( AZ(I) + AZ(I1) ) / 2.
      IF ( AZ(I1) .LT. AZ(I) ) ANG = ANG + 3.141592654
      CALL FUNB( 1, ANG, F, G )
      I1 = NTR + I
      XT(I1) = ANG
 14   SAMP(I1) = F
      NTOT = NTR + NCIR
      PRINT 13, ( I, ( ITR(I,J), J = 1, 3), XT(I), YT(I), SAMP(I),
     , I = 1, NTR )
 13   FORMAT (4I5,2E15.6,15X,E15.6)
      NTR1 = NTR + 1
      DO 304 I = NTR1, NTOT
      X(1) = XC + R * COS( XT(I) )
```

```
              X(2) = XC + R * SIN( XT(I) )
      304     PRINT 11, I, X(1), X(2), XT(I), SAMP(I)
      11      FORMAT (I5,15X,4E15.6)
      C       PICKS OUT THE THREE SMALLEST SAMPLE VALUES AS A GUESS AT WHERE THE
      C       MINIMUM IS, THEN TAKES THE LEAST OF THESE TO LOCATE A MINIMUM (
      C       HOPEFULLY A GLOBAL MIN).
      C       NOTE FUNCTION VALUES HAVE CHANGED SIGN SINCE FMFP GOES FOR MIN.
      C       THUS WE END UP WITH A MAXIMUM WHEN SIGNS ARE REVERSED
              IC = 2 $ IQ = 2
              DO 24 J = 1, 3
              SML = 1.
              DO 18 I = 1, NTOT
              IF ( SAMP(I) .LT. SML ) 16, 18
      16      SML = SAMP(I)
              I1 = I
      18      CONTINUE
              ITRY(J) = I1
              STAMP(J) = SML
              SAMP(I1) = 1.
              IF ( I1 .LE. NTR ) 20, 22
      20      X(1) = XT(I1)
              X(2) = YT(I1)
      C       TEMP. CARDS
              XTMP(J) = XT(I1)
              YTMP(J) = YT(I1)
              XMAX(J) = XTMP(J)
              YMAX(J) = YTMP(J)
      C       END TEMP. CARDS
              CALL FMFP( FUNCT, 2, X, F, G, EST, EPS, LIMIT, IER, H )
              IF ( IER .NE. 3 ) GO TO 230
              IER = 0
              X(1) = ATAN2( X(2), X(1) )
              CALL FMFP( FUNB, 1, X, F, G, EST, EPS, LIMIT, IER, H )
              GO TO 25
      230     CONTINUE
              IF ( IER .NE. 0 ) GO TO 950
      23      XMAX(J) = X(1)
              YMAX(J) = X(2)
              FMAX(J) = F
              GO TO 24
      22      X(1) = XT(I1)
      C       TEMP CARDS
              XTMP(J) = XT(I1)
              YTMP(J) = 99.9
      C       END TEMP CARDS
              CALL FMFP( FUNB, 1, X, F, G, EST, EPS, LIMIT, IER, H )
              IF ( IER .NE. 0 ) GO TO 950
      25      ANG = X(1)
              XMAX(J) = XC + R * COS( ANG )
              YMAX(J) = YC + R * SIN( ANG )
              FMAX(J) = F
      24      CONTINUE
      C       TEMP CARDS
      C       PRINT OUT OF COMPARISON OF ESTIMATED AND FINAL FOR ALL 3 POINTS
              DO 502 J = 1, 3
```

```
        IF ( YTMP(J) .EQ. 99.9 ) 503, 502
503     YTMP(J) = YC + R * SIN( XTMP(J) )
        XTMP(J) = XC + R * COS( XTMP(J) )
502     CONTINUE
        PRINT 504, ( XTMP(J), YTMP(J), STAMP(J), J = 1, 3 )
        PRINT 504, ( XMAX(J), YMAX(J), FMAX(J), J = 1, 3 )
504     FORMAT (3(2F10.5,2X,E12.5))
C       END TEMP CARDS
C       NOW ORDER THESE
        DO 30 K = 1, 3
        SML = 1.
        DO 28 J = 1, 3
        IF ( FMAX(J) .LT. SML ) 26, 28
26      SML = FMAX(J)
        J1 = J
28      CONTINUE
        JTRY(K) = J1
        FM2(K) = SML
        FMAX(J1) = 1.
30      CONTINUE
        IF ( FBEF .GT. 0. ) 32, 34
32      PER = 999.9
        GO TO 36
34      PER = ( FBEF - FM2(1) ) / FBEF
36      DO 38 J = 1, 3
        XA(J) = XMAX( JTRY(J) )
        YA(J) = YMAX( JTRY(J) )
38      FM2(J) = -FM2(J)
        FBEF = -FM2(1)
C       ADDS NEW POINT TO LIST
        X0 = XA(1)
        Y0 = YA(1)
        CALL ADDPT( X0, Y0 )
C       ADDS TO CIRCLE LIST IF NECESSARY
        DIST = SQRT( ( X0 - XC )**2 + ( Y0 - YC )**2 )
        IF ( DIST .GE. ( R - 0.1 ) ) 40, 50
40      ANG = ATAN2( Y0 - YC, X0 - XC )
        IF ( ANG .LT. 0. ) ANG = TPI + ANG
        DO 44 I = 1, NCIR
        I1 = I + 1 $ IF ( I .EQ. NCIR ) I1 = 1
        IF ( AZ(I1) .LT. AZ(I) ) GO TO 42
        IF ( ANG .LT. AZ(I1) .AND. ANG .GT. AZ(I) ) 46, 44
42      IF ( ANG .GT. AZ(I) .AND. ANG .LE. TPI ) GO TO 46
        IF ( ANG .GE. 0. .AND. ANG .LT. AZ(I1) ) GO TO 46
44      CONTINUE
        PRINT 2017, ANG
2017    FORMAT (* FAILED TO FIND NEW CIRCLE POINT FOR ANG = *, F10.5)
        CALL EXIT
46      IA = I1
        I2 = NCIR - I1 + 1
        NCIR = NCIR + 1
        DO 48 I = 1, I2
        I1 = NCIR + 1 - I
        IM = I1 - 1
48      AZ(I1) = AZ(IM)
```

```
      AZ(IA) = ANG
50    CONTINUE
C     PRINTS OUT THE RESULTS OF THIS PASS
      PRINT 2020, NSTN, XA(1), YA(1), FM2(1), PER, ( XA(I), YA(I),
     , FM2(I), I = 2, 3 )
2020  FORMAT (I10,F8.1,F6.1,E11.3,F8.2,F8.1,F6.1,E11.3,F7.1,F6.1,E11.3)
      ISPACE = ISPACE + 1
      IF ( ISPACE .EQ. 5 ) 52, 54
52    ISPACE = 0
      PRINT 2002
54    CONTINUE
C     HAS THE NUMBER ASKED FOR BEEN REACHED
      IF ( NSTN .GE. NSTOP ) GO TO 900
C     TRY AGAIN
      GO TO 8
C     PRINTS REST OF INPUTS
900   IF ( IPRNT .NE. 1 ) GO TO 999
      PRINT 2000
      PRINT 2024
      PRINT 2026, ( J, ITR(J,1), ITR(J,2), ITR(J,3), J = 1, NTR )
2024  FORMAT (8X,*INDEX NUMBERS OF TRIANGLE VERTICES*)
2026  FORMAT (8(1X,I2,1H:,I3,1H,,I3,1H,,I3,1H/))
      PRINT 2000
      PRINT 2028
2028  FORMAT (5X,*PARAMETERS OF THE EMPIRICAL CORRELATION COEFFICENTS*/
     /8X,*COORDINATES OF THE DATA GIRD*)
      PRINT 2030, ( XD(I), I = 1,9 )
2030  FORMAT (* X = *,9(F8.2,1H,))
      PRINT 2032, ( YD(I), I = 1, 9 )
2032  FORMAT (* X = *,9(F8.2,1H,))
2038  FORMAT (1P9E12.4)
      PRINT 2040
2040  FORMAT (*WEIGHT FUNCTION ARRAY IN FORM OF COORDINATE GRID*)
      DO 914 J = 1, 9
      J1 = 10 - J
      PRINT 2038, ( W(I,J1), I = 1, 9 )
914   CONTINUE
      PRINT 2000
      PRINT 2042, XC, YC, R
2042  FORMAT(* OUTER CIRCLE HAS CENTER AT X = *,F7.3,2X,* Y = *,F7.3,2X,
     ,* AND RADIUS = *,F8.3)
      PRINT 2002
      PRINT 2044, NADD, PMIN, EPS, LIMIT
2044  FORMAT (* NUMBER OF POINTS TO BE ADDED = *,I5/* PERCENT CHANGE OF
     *MAXIMUM ERROR OF ESTIMATE BETWEEN ITERATIONS = *,F7.3/
     ** EXPECTED ABSOLUTE ERROR FOR FMFP = *,E10.3/* MAXIMUM NUMBER OF I
     *TERATIONS OF FMFP = *,I5)
      GO TO 999
C     DIAGNOSTICS FROM FMFP
950   N1 = NSTN + 1
      IF ( IER .EQ. 1 ) 952, 954
952   PRINT 2050, N1, LIMIT
2050  FORMAT(* NO CONVERGENCE FOR POINT NO.*,I3,2X,*IN*,I3,2X,
     **ITERATIONS*)
      GO TO 960
```

```
954   IF ( IER .EQ. 2 ) 956, 958
956   PRINT 2052, N1, LIMIT
2052  FORMAT (* NO MAXIMUM. POINT NO. = *,I3,2X,* ITERATIONS = *,I3)
      GO TO 960
958   PRINT 2054, N1, LIMIT
2054  FORMAT(* ERRORS IN GRADIENT.  POINT NO. = *,I3,2X,* ITERATIONS = *
     ,,I3)
960   IF ( I1 .GT. NTR ) GO TO 962
      PRINT 2056, XT(I1), YT(I1), STAMP(J), X(1), X(2), F, G(1), G(2)
2056  FORMAT (* INITIAL VALUES*/2F10.5,E15.5/* FINAL VALUES AND GRADIENT*
     **/2F10.5,3E15.5)
      GO TO 23
962   PRINT 2058, XT(I1), STAMP(J), X(1), F, G(1)
2058  FORMAT(* INITIAL VALUES*/F10.5,E15.5/* FINAL VALUES AND GRADIENT*/
     /F10.5,2E15.5)
      GO TO 25
999   STOP
      END
```

BAST Code Sample Problem

The input required for the BAST code is defined on the following page. Reference to the main subprogram and to the text is helpful for obtaining further detail. A sample input case follows. The case consists of eight fixed stations, the last four of which are boundary points, i.e., their coordinates define the convex region containing the eight stations (data card groups 1-3). A total of 10 triangles cover the region defined by the eight stations (card groups 4 and 5).

The region for which correlation coefficient data is input, and to which additional stations are to be optimally added, is defined by a circle centered at (0,0), with a radius of 30 km (card 6). No input points on the circle are indicated by card 6, so card 7 is not required. In this event, the code will automatically define two points on the perimeter of the circle.

Input card 8 specifies the number of stations to be added, 17, plus program controls for the minimization routine (FMFP) and printing option. Cards 9 and 10 contain the standard set of coordinates for this problem, a 9 X 9 grid with 10 km spacing in both the X and Y directions. Card group 11 lists the proper values of the correlation coefficient matrix, in monotonic decreasing order. The desired significance is obtained with the first 15 values. Card group 12 contains the proper function coefficients, at each point in the 9 X 9 grid, corresponding to the 15 proper values. Last, card group 13 is a 9 X 9 array of weight parameters for the correlation coefficients. The normal situation of equal weighting, all weights being 1.0, is input in this case.

A sample output follows the input. The fixed stations, the stations added on the circle, and then the remaining stations to be added are listed. For each added station the three prospective coordinate pairs of maximum estimate of error (E of E) are listed. The optimization procedures, if successful, varies these coordinates in obtaining the location for which the greatest reduction in estimate of error can be achieved. Optionally, extensive mathematical detail can also be output.

CARD GROUP NUMBER	SYMBOL	DEFINITION	FORMAT
1	NSTN	NUMBER OF STATIONS. MAXIMUM OF 30	I5
2	XS, YS	STATION COORDINATES (KM), FROM ORIGIN. ONE COORDINATE PAIR PER CARD, NSTN CARDS.	2F10.0
3	NBDY IBDY	NUMBER OF BOUNDARY POINTS, MAXIMUM OF 20. INDEX NUMBERS OF BOUNDARY STATIONS, IN CCW ORDER	16I5
4	NTR	NUMBER OF TRIANGLES, MAXIMUM OF 50	I5
5	ITR(I,J)	STATION INDEX NUMBERS OF TRIANGLE VERTICES. CCW ORDER. ONE TRIANGLE PER CARD, NTR CARDS. I=TRIANGLE NUMBER, J=1,2,3 FOR TRIANGLE VERTICES	3I5
6	XC, YC R NCIR	COORDINATES OF CENTER OF CIRCLE CONTAINING REGION RADIUS OF CONTAINING CIRCLE NUMBER OF POINTS ON THE CIRCLE	3F10.0,I5
7	AZ	AZIMUTH (DEG) CCW FROM EAST, OF POINTS LYING ON CIRCLE, TOTAL OF NCIR POINTS. SKIP THIS CARD IF NCIR = 0.	9E8.0
8	NADD PMIN EPS LIMIT IPRNT	NUMBER OF STATIONS TO BE ADDED PERCENT REDUCTION BETWEEN MAXIMUM ERRORS OF ESTIMATE. NO POINTS ADDED AFTER PRECENT REDUCTION GOES BELOW PMIN EXPECTED ABSOLUTE ERROR, SUBROUTINE FMFP MAXIMUM NUMBER OF ITERATIONS IN SUBROUTINE FMFP PRINT OPTION FLAG. 0, OUTPUT STATION COORDINATES ONLY. 1, DETAILED DIAGNOSTICS PRINTED.	I5,F10.0,E10.1,2I5
9 10	XD YD	X-COORDINATES FOR CORRELATION COEFFICIENT INPUT Y-CORRIDNATES FOR CORRELATION COEFFICIENT INPUT DATA. CODE ASSUMES 9 VALUES OF XD AND YD, FOR A 9X9 GRID WITH 10 KM SPACING BETWEEN POINTS.	9F8.0 9F8.0
11	ALAM	PROPER VALUES FOR CORRELATION COEFFICIENT DATA, 15 VALUES.	8F10.0
12	PH(I,J,K)	COEFFICIENTS OF 15 PROPER FUNCTIONS, I = 1, 15, AT GRID POINTS SPECIFIED BY (XD,YD), ORDERED FOR XD VALUES. FOR EACH YD VALUE (I.E., ((PH(I,J,K), K = 1,9), J = 1,9)).	16X,4E16.8/ 5E16.8/ (4E16.8/ 5E16.8)
13	W	WEIGHT FUNCTION TABLE FOR POINTS (XD,YD),	9F8.0

P R O G R A M B A S I I N P U T E X A M P L E

```
***** CARD 1       * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
    8

***** CARD SET 2 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
 0.         0.
 -.5        4.5
 -4.75      -2.5
 -.5        -4.
 4.5        -8.
 4.5        -2.5
 3.5        9.
 -22.       7.5

***** CARD 3       * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
    4    5    6    7    8

***** CARD 4       * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
   10

***** CARD SET 5 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
    8    5    3
    5    4    3
    4    5    6
    4    6    1
    3    4    1
    1    6    2
    2    6    7
    1    2    3
    2    7    8
    2    8    3

***** CARD 6       * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
 0.         0.         30.          0

***** CARD 8       * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
   17 .01        .1        5       1

***** CARDS 9, 10* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
 -40.    -30.    -20.    -10.    0.      10.     20.     30.     40.
 -40.    -30.    -20.    -10.    0.      10.     20.     30.     40.

***** CARD SET 11* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
28.909829920.4335128 9.3925182 5.2604457 2.6616002 2.1965837 1.2201544 1.1834856
  .7342799  .5133393  .3574174  .3318036  .2512566  .2123830  .1081351

***** CARD SET 12* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
PROPER FUNC.  1   2.13492226E-02  2.91629144E-02  4.52380567E-02  6.57938454E-02
  7.82732549E-02  9.72490823E-02  1.08749172E-01  1.09638454E-01  1.11606421E-01
  4.91204939E-03  1.86498594E-02  3.98497936E-02  6.74289853E-02
  9.71762028E-02  1.11502420E-01  1.13954070E-01  1.15293137E-01  1.27428301E-01
 -1.13289246E-02 -1.36877176E-03  2.56043344E-02  6.20960851E-02
  9.54481463E-02  1.31071366E-01  1.42071804E-01  1.41159325E-01  1.39988929E-01
 -5.52873886E-02 -5.24850443E-02 -2.57480476E-03  2.58566478E-02
```

2.39534203E-02	1.46224196E-01	1.63241569E-01	1.61389957E-01	1.58256587E-01
-1.12859460E-01	-1.47511622E-01	-8.83894338E-02	-6.44476884E-02	
-4.40726114E-02	9.97209066E-02	8.17998193E-02	1.57191973E-01	1.59515717E-01
-1.53931315E-01	-1.68599575E-01	-1.64035302E-01	-1.49250796E-01	
-1.19492211E-01	2.58123666E-03	8.87039067E-02	1.48033644E-01	1.42279105E-01
-1.63982476E-01	-1.69374417E-01	-1.68135129E-01	-1.35013178E-01	
-1.11995777E-01	-7.04099538E-02	9.23905138E-03	1.09743374E-01	1.22256099E-01
-1.66174576E-01	-1.66497949E-01	-1.63204792E-01	-1.38751474E-01	
-9.32753081E-02	-3.45392130E-02	4.27811924E-02	9.06458659E-02	1.02960437E-01
-1.57249352E-01	-1.53433349E-01	-1.52328082E-01	-1.40251636E-01	
-8.98516291E-02	-3.51570809E-02	2.53935571E-02	6.86277712E-02	8.67620603E-02
PROPER FUNC. 2	1.33955398E-01	1.36338386E-01	1.37039247E-01	1.27161225E-01
1.21418473E-01	1.10422289E-01	1.01413802E-01	8.86563383E-02	8.22119671E-02
1.51106930E-01	1.52675177E-01	1.46852656E-01	1.50468676E-01	
1.30656144E-01	1.13978242E-01	9.50011805E-02	8.10908563E-02	6.73524410E-02
1.47567607E-01	1.57509329E-01	1.74157948E-01	1.79579713E-01	
1.45989865E-01	9.80388147E-02	6.00147775E-02	4.12175317E-02	3.09312749E-02
1.51173858E-01	1.57258546E-01	1.69148551E-01	1.53953889E-01	
1.17485730E-01	5.14048116E-03	-2.31934989E-02	-2.30572652E-02	-2.23384295E-02
1.25469179E-01	9.37120367E-02	3.20450627E-02	6.67139464E-03	
2.26306189E-02	-7.75524176E-03	-6.05951654E-02	-7.13192021E-02	-6.36194797E-02
9.34450852E-02	6.98945000E-02	2.78229525E-02	-1.13676616E-02	
-2.58696663E-02	-6.47530978E-02	-1.19629619E-01	-1.10578396E-01	-1.07124958E-01
6.98924213E-02	6.31941974E-02	4.45936300E-02	-9.25535914E-03	
-9.74253488E-02	-1.63026999E-01	-1.86228964E-01	-1.53746234E-01	-1.41199446E-01
3.33584700E-02	1.83596197E-02	-9.70309859E-03	-7.07914914E-02	
-1.50397206E-01	-1.79858970E-01	-1.67616599E-01	-1.58898005E-01	-1.50469955E-01
4.44365101E-03	-7.85175733E-03	-2.67778481E-02	-5.40897611E-02	
-1.13950357E-01	-1.56834984E-01	-1.56827710E-01	-1.57678765E-01	-1.56714116E-01
PROPER FUNC. 3	1.62889284E-01	1.43874803E-01	8.73750232E-02	-7.64597027E-03
-6.27379823E-02	-1.36572510E-01	-1.78252732E-01	-1.94920797E-01	-1.98658066E-01
1.99934166E-01	1.77069179E-01	1.15742038E-01	2.38448945E-02	
-9.34816879E-02	-1.54679696E-01	-1.83441800E-01	-1.94498252E-01	-1.92617138E-01
1.97881192E-01	1.87439340E-01	1.27883568E-01	4.58400876E-02	
-6.55747051E-02	-1.30100706E-01	-1.50127674E-01	-1.50566801E-01	-1.47045048E-01
1.56034050E-01	1.32088468E-01	3.47956243E-02	-8.66192340E-04	
-8.68863409E-02	2.13103161E-02	-2.42138136E-02	-5.56781904E-02	-7.05904300E-02
6.55262659E-02	1.30054623E-02	1.28710526E-01	1.65188643E-01	
8.82847861E-02	8.22219292E-02	-4.85913439E-02	-2.45316707E-03	-1.44805651E-02
-1.35174779E-02	-3.88702948E-02	4.79845843E-03	4.23319640E-02	
3.27607091E-02	4.85650527E-02	3.37205499E-02	3.50098277E-02	4.78872301E-02
-5.10654671E-02	-4.81305873E-02	-4.43991337E-02	-1.24465882E-01	
-1.11838666E-01	-7.54070277E-02	-3.74477974E-02	8.32360103E-02	1.05117086E-01
-1.05666130E-01	-1.18037865E-01	-1.18175254E-01	-1.01812706E-01	
-4.88358662E-02	-2.15548243E-02	4.82764124E-02	1.13766332E-01	1.29985820E-01
-1.42545836E-01	-1.49775334E-01	-1.45369945E-01	-1.31775470E-01	
-6.95018204E-02	-1.04295290E-03	7.35857366E-02	1.22166653E-01	1.35492323E-01
PROPER FUNC. 4	-1.68681911E-01	-1.97910948E-01	-2.31190148E-01	-2.31451361E-01
-2.05650811E-01	-1.39126975E-01	-8.30775329E-02	-3.79634778E-02	-1.59459740E-02
-7.41059496E-02	-1.37474800E-01	-2.01802683E-01	-2.13780038E-01	
-1.73541871E-01	-1.08525098E-01	-5.30190174E-02	-1.01876163E-02	3.57769469E-02
1.83252732E-02	-1.78439623E-02	-5.26521920E-02	-4.87750871E-02	
-1.11034185E-01	-5.93995993E-02	3.88827128E-02	8.81884828E-02	1.11863306E-01
1.33484552E-01	1.08637939E-01	1.22795588E-01	1.75173792E-01	

4.79127889E-02	-4.19077722E-02	5.58751303E-02	1.12053565E-01	1.36665303E-01
1.61383497E-01	5.04633799E-02	5.79347590E-02	9.59673450E-02	
5.00561823E-02	1.12531094E-01	1.23866973E-01	8.62527037E-02	1.26642007E-01
9.63918369E-02	1.19175690E-02	-5.81447162E-02	-6.05960528E-02	
-8.84098203E-02	2.05583161E-02	-1.01731699E-01	2.05325896E-02	9.74223065E-02
8.39574686E-02	3.80832003E-02	-3.73202978E-02	-4.20683343E-02	
2.87043719E-02	4.47520812E-02	-7.76904342E-02	-1.23762169E-02	4.02994477E-02
5.43366953E-02	1.52670507E-02	-4.76274290E-02	-1.17651010E-01	
-1.09827667E-01	-1.53294038E-01	-1.44077746E-01	-5.04370117E-02	1.75831255E-05
-2.54485856E-03	-3.44833286E-02	-8.08796589E-02	-1.43434473E-01	
-2.31335639E-01	-2.46864654E-01	-2.02407823E-01	-1.10036412E-01	-5.79367640E-02
PROPER FUNC. 5	-1.59884016E-01	-1.69283267E-01	-1.44079465E-01	-4.48672137E-02
2.94705560E-02	1.18602848E-01	1.61987401E-01	1.80939852E-01	1.72210495E-01
-4.96574543E-02	-1.02309609E-01	-1.26638826E-01	-3.28048522E-02	
5.71023814E-02	1.14281782E-01	1.43031982E-01	1.44502398E-01	7.53392041E-02
5.29385061E-02	3.98242058E-02	8.07079618E-02	1.04635854E-01	
2.46266909E-02	6.71795618E-03	-6.44270792E-02	-9.76590493E-02	-1.09465773E-01
1.04562477E-01	9.28737775E-02	1.74291039E-01	-1.57958309E-02	
-2.09007517E-01	5.00997986E-03	-8.83947560E-02	-1.26137706E-01	-1.36188963E-01
6.88108760E-02	3.28550326E-03	2.00073625E-01	5.73086065E-02	
-9.81585790E-02	2.09153813E-01	1.94290067E-02	-2.82802792E-02	-8.83940150E-02
-3.79426854E-02	-4.84231294E-02	1.06278643E-01	2.06090057E-01	
2.53804551E-01	3.04271110E-01	-2.15774472E-02	5.55912273E-02	-3.17293405E-02
-5.16735579E-02	-4.06525149E-02	1.93759268E-02	2.30249159E-01	
1.07424418E-01	2.25169371E-02	-9.99289954E-02	3.67194845E-02	5.42138412E-03
-6.49296552E-02	-6.95967777E-02	-2.44950367E-02	6.40245282E-02	
-5.39149076E-03	-1.07521240E-01	-5.86725843E-02	3.98777725E-02	3.07191213E-02
-1.24685118E-01	-1.30125628E-01	-8.52769331E-02	-4.88553922E-02	
2.25288490E-02	3.18106940E-02	1.16020330E-01	1.62008794E-01	1.27163154E-01
PROPER FUNC. 6	-2.17056352E-02	3.46031725E-02	1.31951755E-01	2.30392728E-01
2.32072364E-01	1.42451769E-01	4.79775388E-02	-1.93042348E-02	-4.93870213E-02
-1.41854046E-01	-5.68104865E-02	5.78099809E-02	1.42575973E-01	
1.34712190E-01	5.01897409E-02	-1.11974066E-02	-5.06651130E-02	-8.56842329E-02
-1.99458596E-01	-1.52454734E-01	-8.97562707E-02	-1.03363937E-01	
-2.99350679E-02	-1.26591154E-02	-4.78178704E-02	-7.82994480E-02	-1.01727867E-01
-6.64851468E-02	-2.83140748E-02	2.07228736E-03	3.32411062E-02	
4.66256801E-02	-9.41750679E-03	1.16133254E-01	4.33274533E-02	-6.69461154E-03
8.67568695E-02	2.36755605E-02	2.61291262E-01	3.56593238E-01	
3.48734889E-01	7.94972357E-02	3.53452308E-01	1.33120088E-01	4.54864518E-02
7.23906777E-02	-4.07019471E-02	-3.13582667E-02	9.03230076E-02	
1.05074735E-01	1.53376502E-02	-1.25392212E-01	2.05791661E-02	4.84856476E-02
4.03795814E-02	-5.15079495E-02	-8.09903577E-02	-5.89128916E-02	
-9.44335972E-02	2.06959990E-02	1.05395273E-02	-9.16775813E-03	6.75449607E-02
4.83501856E-02	1.65552303E-02	-1.46884598E-03	6.25410204E-02	
1.41463880E-01	1.65556502E-02	-1.58885894E-01	-2.61305459E-02	8.04872955E-02
9.49936208E-02	7.28338007E-02	2.88419165E-02	-1.47040130E-02	
-3.55698620E-02	-5.07051690E-02	-1.08357975E-01	1.39613657E-03	6.85158455E-02
PROPER FUNC. 7	-1.00427459E-02	4.49354654E-02	1.21864921E-01	1.46079962E-01
1.02962321E-01	1.61914444E-02	-4.06584780E-02	-7.98054349E-02	-8.81331683E-02
-1.15002120E-01	-4.34455398E-02	2.74285861E-02	3.05759671E-02	
4.55958905E-02	1.37214912E-03	-3.98076201E-02	-6.67310785E-02	-4.95471852E-02
-1.36215949E-01	-1.03513198E-01	-1.08180411E-01	-1.20526417E-01	
-9.29075845E-02	4.38192476E-02	4.82979529E-02	4.00501930E-02	2.58478857E-02
1.40761748E-01	2.09762566E-01	1.88815984E-01	2.24656696E-01	

-2.82799542E-02	1.44426703E-01	-1.69791258E-02	-1.64141748E-01	-1.66373692E-01
-3.73925194E-02	-2.17659356E-01	-2.19994153E-01	7.28788579E-02	
3.12483205E-01	-5.40862427E-02	4.35953040E-02	5.36722003E-02	-1.41141812E-02
9.88434897E-02	-3.70632779E-02	1.02074529E-01	1.19800649E-01	
8.18345205E-02	-3.78424259E-01	2.19916306E-01	6.51355672E-02	2.08636033E-01
1.29770559E-01	-7.39057512E-02	-1.44021211E-01	1.64537456E-01	
1.97011792E-01	5.54416180E-02	-3.37315242E-02	-1.42452250E-01	7.24852736E-02
1.08734877E-01	-3.87271022E-02	-1.88156165E-01	1.31332697E-01	
-1.91314695E-02	5.99714208E-03	1.62038193E-01	-1.64280250E-01	1.36424767E-02
6.44854496E-02	-3.09197076E-02	-1.81930745E-01	-1.33654454E-01	
1.88396135E-02	3.36572479E-02	1.18649374E-01	-1.64353495E-01	-3.74525670E-02
PROPER FUNC. 11	8.25683898E-02	4.67372112E-02	-8.76199945E-02	1.25526006E-01
1.33987406E-01	-6.68833245E-02	-1.35039088E-01	-1.43694424E-01	-1.20920705E-01
1.26697376E-01	-1.12431427E-02	-2.77400548E-01	-1.33852487E-02	
2.37635991E-01	1.52806118E-01	-5.21807478E-02	-1.01401099E-01	-7.47477263E-02
1.48579748E-01	6.92377429E-02	-5.96156752E-02	4.96647756E-03	
5.08409717E-02	2.09975371E-01	2.92726487E-02	-8.49631194E-02	-3.07121629E-02
3.18742356E-02	-1.05726574E-02	3.58266321E-02	-4.28013854E-02	
2.47013442E-02	-5.85949816E-02	1.18333354E-01	-8.05036486E-02	-3.10698470E-02
6.10542273E-02	1.53493866E-01	-5.34112930E-02	-2.53752056E-01	
1.42651270E-01	1.54643319E-01	1.08657167E-01	-4.37308901E-03	9.26963968E-02
-4.72893111E-02	1.33905516E-01	-1.15508569E-01	-3.75528003E-01	
2.27701383E-01	1.14668737E-01	5.79991186E-02	-4.01467295E-02	1.09424660E-01
-1.28835119E-01	-4.42185237E-02	1.67100041E-02	1.40837678E-01	
2.12842445E-01	1.34854949E-01	1.06649602E-01	5.51984419E-02	-4.62945766E-02
2.47074943E-02	7.68438087E-02	-3.12505648E-02	1.28446654E-02	
-2.28151484E-02	-1.58924964E-03	-4.41424206E-02	5.02462214E-02	-1.11362691E-01
3.45244999E-02	5.41046566E-02	-4.69431318E-02	-7.30684972E-02	
3.70660780E-02	-2.17467732E-02	-5.57934482E-02	5.15359284E-02	-7.15978853E-02
PROPER FUNC. 12	1.66360294E-01	-4.83050264E-03	-2.67814981E-01	-1.58706616E-01
3.24804198E-02	1.96620073E-02	-4.74681887E-02	-1.80044029E-02	1.69668501E-02
5.10610381E-02	-1.11173808E-02	2.47470731E-02	2.67088595E-01	
1.24250744E-02	-1.58939523E-01	-1.80207705E-01	-3.30801022E-02	7.92209071E-02
-1.10554399E-01	-2.15871839E-01	2.09581237E-02	9.80612335E-02	
1.03066976E-01	-1.21662529E-01	-9.61946388E-02	7.42226942E-02	1.81907677E-01
3.16872594E-02	-3.40329735E-02	1.33714670E-01	1.60158042E-01	
8.53692248E-02	2.20493176E-01	-1.05652201E-01	-3.00225002E-02	5.60393192E-02
6.16720501E-02	-1.62056975E-02	-2.47607814E-01	-1.34431552E-01	
2.20869430E-01	1.02347831E-01	-1.23351855E-03	7.90950935E-02	7.93575740E-02
8.45785730E-03	2.37538612E-02	-4.47791703E-02	4.26092091E-02	
2.35151781E-01	1.61171751E-01	-5.89546680E-02	7.94554863E-02	-7.66805927E-02
-1.02779290E-03	1.11951373E-01	2.09907670E-02	8.66437358E-02	
-2.25835953E-01	-1.80641844E-01	-7.18872245E-02	-1.08801646E-02	-7.26753405E-02
-7.42316933E-02	-2.13323135E-04	-1.22080644E-02	1.56270111E-01	
4.23024553E-02	9.61063621E-04	1.38365817E-01	1.29292915E-01	2.20737965E-02
5.39579118E-02	1.04121059E-01	6.00743126E-02	6.36930130E-02	
2.58440917E-02	-5.95404601E-02	3.37226852E-02	7.05711687E-02	2.78192958E-02
PROPER FUNC. 13	1.21001596E-02	1.01169555E-01	8.90512828E-02	-9.18772556E-02
-7.75100209E-02	1.59254164E-01	1.45128446E-01	4.11061018E-02	-2.53666764E-02
-5.30325811E-02	1.17056649E-01	1.36011433E-01	-5.29835447E-02	
-1.81153422E-01	-3.76684478E-02	1.16674789E-01	4.09455492E-02	-3.73496478E-02
-7.33847355E-02	7.10790786E-02	2.53084127E-02	-3.43141800E-02	
-6.60771824E-02	-1.19436210E-01	1.24555111E-01	6.57158294E-02	-4.49639523E-02

-2.33119562E-01	7.49706025E-02	-3.10036466E-02	-7.38588098E-02	-7.36717949E-02
8.63910562E-02	-2.24328346E-02	-9.15744706E-02	-1.08374731E-01	
-2.63571796E-01	2.45782225E-01	8.61775769E-02	-9.09715457E-02	-1.15782423E-01
-1.09626299E-01	-2.87961886E-02	5.40191011E-02	-4.46324273E-02	
-1.88591191E-01	-3.55138748E-01	-2.64213382E-03	3.76798768E-03	-7.85767639E-02
-1.13790812E-01	-2.22969691E-02	7.69425174E-02	4.34096829E-02	
6.83412218E-02	4.62434871E-02	-7.82985369E-02	7.78738087E-02	1.95017073E-02
-2.44341837E-02	2.09334029E-02	1.14629427E-01	1.44210041E-01	
3.41588728E-02	-1.37868063E-01	1.16805036E-01	1.58239824E-01	7.81928067E-02
1.81450890E-02	5.15112172E-02	1.06476666E-01	6.68666889E-02	
-1.44408245E-01	-2.15220287E-01	4.24063544E-02	1.02688843E-01	8.28962240E-02
PROPER FUNC. 8	-2.41039756E-01	-2.15039539E-01	-7.58274555E-02	1.50842204E-01
2.19293326E-01	1.57219292E-01	5.08999623E-02	-1.68924878E-02	-4.27155502E-02
-1.01529704E-01	-1.48714753E-01	-1.15499999E-01	5.61374056E-02	
7.63922850E-02	5.61050410E-03	-3.79478146E-02	-5.85598554E-02	-7.92641030E-02
4.06165171E-02	4.99057414E-02	1.21392317E-01	2.94802764E-02	
-1.76022413E-01	-1.09001458E-01	-7.41219842E-02	-7.00720379E-02	-7.69235880E-02
3.73390348E-02	3.23056805E-02	-2.06893589E-02	-1.16952558E-01	
-2.65477559E-01	-2.70715035E-01	2.29711958E-02	1.18244430E-01	7.55327438E-02
-6.57878989E-02	-3.05731488E-02	-1.84909231E-01	-1.15494850E-01	
1.86058399E-01	-1.10037346E-02	-1.09974687E-01	5.38506599E-02	7.18429633E-02
-2.92134157E-02	5.45035960E-02	-7.12299976E-02	-9.87144237E-03	
5.30631914E-02	-1.90038286E-01	-2.41381059E-02	-4.97904783E-02	-5.32608839E-02
3.32088674E-02	5.02251035E-02	-4.41020220E-02	-1.07982474E-01	
-1.81341067E-01	-2.10244087E-01	-1.39192232E-01	-8.59953171E-02	-5.92509549E-02
1.09953670E-01	9.32370442E-02	-3.31457204E-02	-1.57013208E-01	
-4.30207760E-02	-1.14897422E-02	1.32215385E-01	-1.38915543E-02	-5.26872546E-02
3.40049886E-02	1.49254773E-02	-6.21015472E-02	-1.12170645E-01	
-1.36705281E-01	7.26608846E-02	2.01357978E-01	3.03754746E-02	-4.35972384E-02
PROPER FUNC. 9	2.12001231E-01	1.83879709E-01	2.82095029E-02	-1.69783053E-01
-2.07738310E-01	-1.02879786E-01	3.01859057E-02	1.12394611E-01	1.16637655E-01
3.26869989E-02	9.05219960E-02	6.96741729E-02	-5.89582951E-02	
-2.81299804E-02	5.90876378E-02	9.88532711E-02	9.52506936E-02	7.95045234E-03
-1.14254261E-01	-1.28714919E-01	-1.74573395E-01	-5.73760548E-02	
1.60890183E-01	2.77820621E-04	-1.22887469E-01	-1.86162901E-01	-1.90385064E-01
3.98437119E-02	8.19727140E-02	1.82117678E-01	-3.22036246E-02	
-2.49827805E-01	-1.90953110E-01	8.16252814E-02	1.10289532E-01	8.34999555E-02
4.01506724E-02	5.16372637E-02	6.34156420E-02	-1.16985973E-01	
-1.44157838E-01	-3.15893242E-02	1.29352613E-01	6.55147188E-02	1.43773961E-01
-5.45049151E-03	1.59182203E-02	-1.13131394E-01	-2.55872632E-02	
1.30046870E-01	1.26838898E-02	2.15014920E-01	3.97955814E-03	1.04355686E-01
1.77772293E-03	6.97718269E-02	-2.69780237E-02	-1.75489527E-01	
-1.46916036E-01	-8.75714845E-02	1.98914911E-03	-1.73697098E-02	-1.50286357E-02
1.56630860E-02	6.46863539E-02	1.93379237E-02	-4.82005994E-02	
1.36130580E-01	-9.93326779E-02	-8.59225470E-02	-1.03637936E-01	-1.29696080E-01
1.20347300E-01	1.52708981E-01	9.40427388E-02	-8.00150625E-03	
-1.53732209E-01	4.86043964E-02	1.89917113E-01	8.14907781E-02	-4.41473764E-02
PROPER FUNC. 10	5.19245337E-02	5.25445878E-02	-8.53348695E-03	-6.10131986E-02
-7.86148557E-02	-5.68516780E-02	1.61590240E-03	4.37464475E-02	4.71312407E-02
1.46658645E-02	2.61797014E-02	-2.76435206E-02	-5.56409019E-02	
3.47263658E-03	4.08572277E-02	4.86359849E-02	5.75054072E-02	2.71922774E-02
-6.06995400E-03	-7.70839837E-03	-5.76430355E-02	-8.00131994E-03	
4.81392726E-02	7.35752060E-02	-2.02514430E-03	-3.75430098E-02	-4.83435257E-02
6.24481046E-02	7.72921194E-02	1.08477750E-01	4.67961410E-02	

-4.05590052E-02	1.13100101E-01	-9.36823091E-02	-1.53302877E-02	
-5.30618510E-03	-2.75630009E-02	3.45953949E-02	-6.88182444E-02	-1.57768380E-01
-6.07213021E-02	1.72880856E-01	-1.60084127E-01	-1.28445577E-01	
1.20188588E-01	1.32251558E-01	9.81075947E-02	1.13771869E-01	-3.15872960E-02
-1.50699222E-01	4.46990373E-02	9.21329765E-02	1.02980647E-01	
-1.32679503E-01	1.50578742E-01	-2.66728730E-01	-2.45049420E-01	1.27557639E-01
-3.07615636E-02	1.72212085E-01	-2.74877127E-01	-2.33524471E-02	
5.27634538E-02	9.36939798E-02	7.96877317E-02	8.29051562E-02	-8.35750966E-02
1.63238222E-02	1.36863037E-01	-1.71728320E-01	2.99866893E-02	
-2.08352659E-02	-1.92075460E-02	7.34410670E-02	1.80221424E-01	-1.59499634E-01
8.18360227E-02	1.37894068E-01	-1.25761583E-01	-1.56662924E-01	
4.67046101E-02	-4.75124062E-02	9.48417994E-03	1.70286337E-01	-7.88647091E-02
PROPER FUNC. 14	1.37521337E-01	1.23282483E-01	-2.98964997E-01	-2.32165849E-01
2.00213242E-02	3.78337442E-01	1.75090830E-01	-9.24308858E-02	-1.90807610E-01
5.28417312E-02	1.07317980E-01	-2.32956640E-01	-4.14540917E-02	
-7.38812440E-02	7.61643881E-02	1.23057689E-01	-9.07752901E-02	-1.71256741E-01
5.81050621E-03	9.34289127E-02	-1.95853042E-02	-7.20719232E-03	
-5.48997783E-02	-9.81769533E-02	1.81937546E-01	-2.72677745E-02	-8.81564734E-02
-1.41158515E-02	1.23551373E-01	5.26662167E-02	-4.39765268E-02	
3.33668940E-02	6.44223548E-02	1.28538508E-01	1.53829556E-02	-1.90309183E-02
-1.64419620E-01	-5.12842159E-02	1.79275259E-01	-1.71821853E-01	
3.30006154E-02	-1.42981904E-01	1.99650796E-01	-6.85852177E-02	1.70975227E-02
-5.46608587E-02	-1.46062995E-01	2.24191385E-01	-1.35044622E-01	
-3.14438469E-02	-7.45787842E-02	-2.82859856E-02	7.22864071E-02	-1.10639738E-01
7.03747368E-02	-2.53569499E-02	1.68952018E-03	3.96621159E-02	
-1.13225561E-01	-7.75720922E-02	-7.80894012E-02	-3.99834300E-02	2.43128710E-02
9.43947797E-03	-3.96437435E-02	8.01659213E-02	3.18964091E-03	
1.56216338E-02	4.15892431E-02	2.78995679E-02	-4.04287318E-02	1.02598476E-01
-2.55055303E-02	-4.57189754E-02	8.07730036E-02	1.03856222E-01	
-2.37234579E-02	1.28881170E-02	-1.55241502E-02	-8.84568572E-02	3.96040558E-02
PROPER FUNC. 15	5.21964872E-02	-2.10544996E-01	5.22399035E-02	7.81234912E-03
-1.75163969E-02	2.12823579E-01	2.02464082E-02	-2.37720429E-01	-2.50276224E-02
2.14161628E-01	-2.65948876E-01	1.46343069E-01	1.32038362E-02	
-6.28205997E-02	1.54014545E-01	9.86402888E-02	-3.11376437E-01	7.80058348E-02
1.63966818E-01	-2.56435664E-01	3.27300924E-02	5.34787910E-02	
3.35785569E-02	9.51913983E-02	7.40708063E-02	-2.13450843E-01	1.30892220E-01
1.72608575E-01	-1.20308773E-01	-1.76011352E-01	1.18513980E-01	
-5.70547759E-02	8.54459538E-03	3.53012724E-02	-1.72881538E-01	1.10039234E-01
4.02182846E-02	1.70372901E-01	-8.97916020E-02	-1.23126365E-02	
-1.59906235E-01	-2.22630286E-01	4.56009995E-02	-1.35820343E-02	4.02632256E-02
-5.33178438E-02	-5.81248071E-02	1.23019119E-01	1.71200229E-01	
1.05015452E-02	6.13856509E-02	9.20080344E-02	6.60724495E-02	8.59649248E-02
3.16619165E-02	-4.16477383E-04	-1.61299361E-01	-5.91999293E-02	
-2.14736746E-03	4.51422908E-02	-3.23173323E-02	-6.33444631E-03	2.63348780E-02
1.14704035E-02	4.14370457E-03	-6.44912525E-02	3.51965630E-02	
7.58568479E-02	-7.18001541E-02	-4.95195973E-02	-3.29844789E-02	-1.40607685E-02
4.04554717E-02	3.63465848E-02	-2.95559179E-02	-2.85140529E-02	
-5.22785394E-02	-5.42871150E-02	-4.73866483E-03	6.39148208E-03	-1.60590409E-03

***** CARD SET 13* *

1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

A. COORDINATES OF INITIAL STATIONS

SERIAL NUMBER	COORDINATES X	Y
1	0.0	0.0
2	-.5	4.5
3	-4.9	-2.5
4	-.5	-4.0
5	4.5	-3.0
6	4.5	-2.5
7	3.5	9.0
8	-22.0	7.5

B. COORDINATES OF CIRCLE POINTS ADDED

9	29.0	-7.6
10	-29.0	7.6

C. COORDINATES OF ADDED POINTS AND SUPPLEMENTARY DATA

SERIAL NUMBER	COORDINATES X	Y	E OF E BEFORE	PERCENT REDUCT.	ALTERNATIVES REJECTED X	Y	E OF E	X	Y	E OF E
11	-7.6	-29.0	.558E+00	999.90	7.6	29.0	.558E+00	-9.8	-1.0	.558E+00
12	-25.9	-15.1	.232E+00	.58	7.6	29.0	.232E+00	15.1	-25.9	.232E+00
13	7.6	29.0	.482E-01	.79	15.1	-25.9	.482E-01	-18.1	-23.9	.482E-01
14	13.4	10.1	.481E-01	.00	-18.1	-23.9	.481E-01	12.3	-.4	.481E-01
15	-18.1	-23.9	.453E-01	.05	-12.8	-15.6	.453E-01	7.1	5.5	.453E-01
16	-12.8	-15.6	.419E-01	.08	-9.1	3.2	.419E-01	7.1	5.5	.419E-01
17	-9.1	3.2	.411E-01	.02	7.1	5.5	.411E-01	-15.1	25.9	.411E-01
18	7.1	5.5	.405E-01	.01	-11.9	2.7	.405E-01	7.5	-.1	.405E-01
19	-11.9	2.7	.405E-01	.00	-15.1	25.9	.405E-01	8.0	8.2	.405E-01
20	8.0	8.2	.399E-01	.01	-15.1	25.9	.399E-01	-29.7	-4.1	.399E-01
21	-15.1	25.9	.398E-01	.00	-29.7	-4.1	.398E-01	15.6	-1.8	.398E-01
22	-1.3	21.3	.395E-01	.01	-13.6	14.2	.395E-01	-29.7	-4.1	.395E-01
23	-6.6	12.6	.394E-01	.00	-29.7	-4.1	.394E-01	8.6	-14.9	.394E-01
24	-29.7	-4.1	.387E-01	.02	8.6	-14.9	.387E-01	15.6	-1.8	.387E-01
25	8.6	-14.9	.140E-01	.04	-1.5	14.3	.140E-01	-4.1	29.7	.140E-01

B. Subroutine CIRCUM

Subroutine CIRCUM is used to locate points on the circular boundary if not more than one is already located there. If one point is already on the boundary the subroutine locates the second point on the boundary (line 35) at the point of maximum residual error. The search is started at a point opposite the point already located on the boundary. If no point has been located on the boundary, the "center of gravity" of the already located interior stations is found. The first point is then located tentatively at the point in which the line through the "center of gravity" of already located points and the center of the circle meets the circumference of the circular boundary. A minimum of the residual variances is then sought from that starting point.

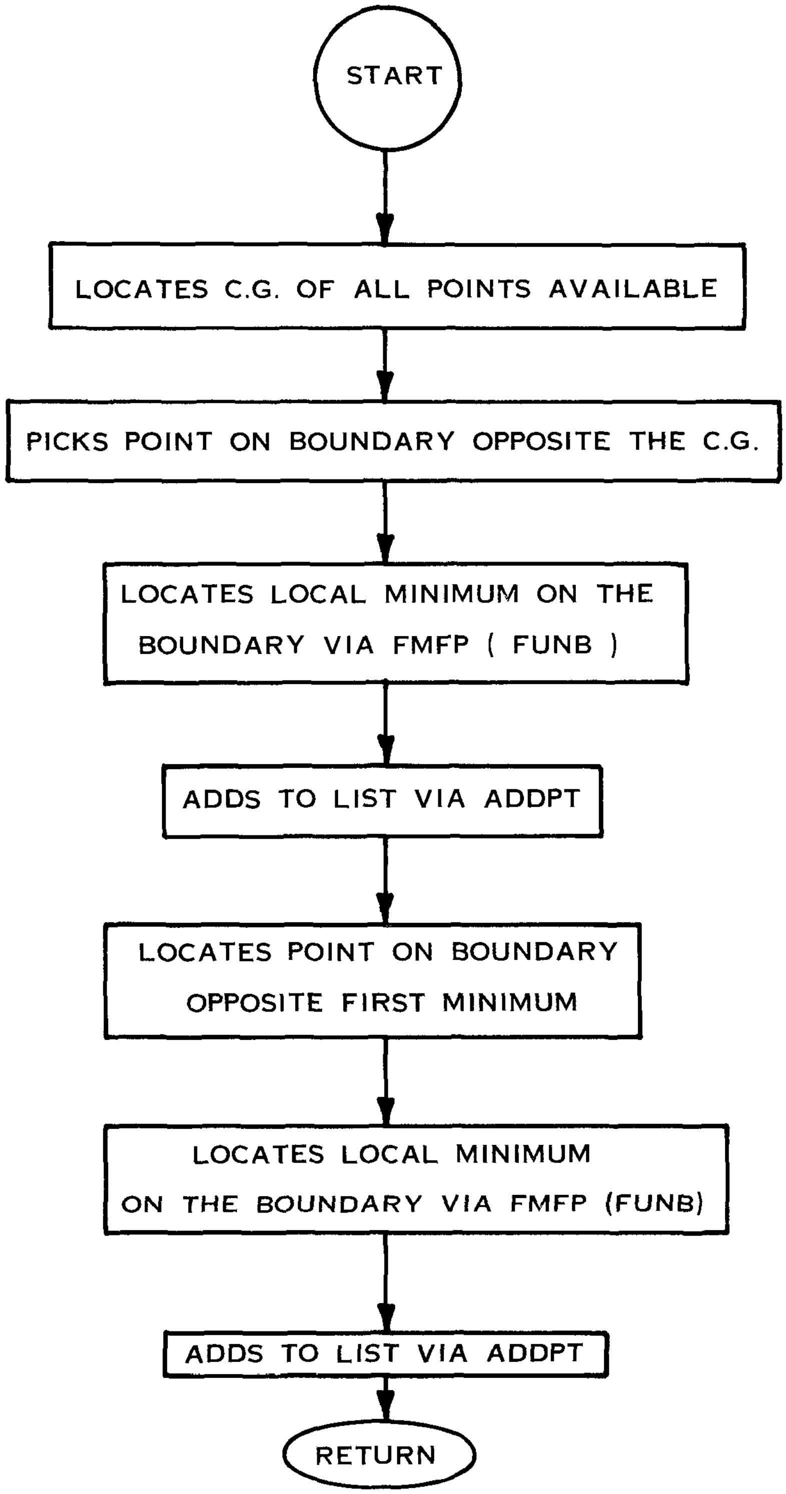


FIGURE V-2

ABBREVIATED FLOW CHART FOR SUBROUTINE CIRCUM.

```
*DECK CIRCUM
      SUBROUTINE CIRCUM( XC, YC, R, I )
C     LOCATES POINTS ON THE CIRCUMFERENCE IF NOT MORE THAN ONE IS
C     ALREADY THERE.  ( XC, YC ) IS THE CIRCLE CENTER, R = RADIUS
      COMMON / BLK1 / NSTN, NBDY, NTR, XS(30), YS(30), ITR(50,3)
      COMMON / BLK3 / NCIR, AZ(20)
      COMMON / BLK4 / EST, EPS, LIMIT, IER, F
      COMMON / LINKS / ERR1, X0(60), DIFF(60)
      EXTERNAL FUNB
      DATA PI, TPI / 3.1415927, 6.2831853 /
      IF ( I .EQ. 1 ) GO TO 20
C     FINE CG OF POINTS AVAILABLE
      ANO = NSTN
      SUM1 = SUM2 = 0.
      DO 10 I = 1, NSTN
      SUM1 = SUM1 + XS(I)
10    SUM2 = SUM2 + YS(I)
      XCG = SUM1 / ANO
      YCG = SUM2 / ANO
      DIST = SQRT( ( XCG - XC )**2 + ( YCG - YC )**2 )
C     SELECTS POINT OPPOSITE CG OR WEST
      IF ( DIST .LE. 1. ) 12, 14
12    X = XC - R
      Y = YC
      GO TO 16
14    X = XC - ( XCG - XC ) * R / DIST
      Y = YC - ( YCG - YC ) * R / DIST
16    ANG = ATAN2( Y-YC, X-XC )
18    CALL FMFP( FUNB, 1, ANG, F, G, EST, EPS, LIMIT, IER, H )
      X0 = XC + R * COS( ANG )
      Y0 = YC + R * SIN( ANG )
      CALL ADDPT( X0, Y0 )
      IF ( ANG .LT. 0. ) ANG = ANG + TPI
      NCIR = 1
      AZ(1) = ANG
C     PUTS ANOTHER POINT ON CIRCLE
20    ANG = AZ(1) + PI
      CALL FMFP( FUNB, 1, ANG, F, G, EST, EPS, LIMIT, IER, H )
      IF ( ANG .GT. TPI ) ANG = ANG - TPI
      AZ(2) = ANG
      NCIR = 2
      X0 = XC + R * COS( ANG )
      Y0 = YC + R * SIN( ANG )
      CALL ADDPT( X0, Y0 )
      RETURN
      END
```

C. Subroutine FMFP

Subroutine FMFP is a "standard" method of finding the minimum of a function of several independent variables and was extracted from the IBM System/360 Scientific Subroutine Package (360-CM-03X) Version III, Programmer's Manual. The details of the minimization procedure are described in the above and also in R. Fletcher and M.J.D. Powell, "A Rapidly Convergent Descent Method of Minimization", Computer Journal, Vol. 6, No. 2, 1963, pp. 163-168. The subroutine as listed in the first reference above has been modified at lines 195-200. In our application the standard FMFP iterative procedure did not meet the convergence criterion. A loop was added to check the magnitude of $t-\alpha h_i/x_i$ and if it was found to be less than ε for all i, then a branchout of the iteration was made by transferring to statement 36.

```
*DECK FMFP
      SUBROUTINE FMFP(FUNCT,N,X,F,G,EST,EPS,LIMIT,IER,H)
C
C     PURPOSE
C        TO FIND A LOCAL MINIMUM OF A FUNCTION OF SEVERAL VARIABLES
C        BY THE METHOD OF FLETCHER AND POWELL
C
C     USAGE
C        CALL FMFP(FUNCT,N,X,F,G,EST,EPS,LIMIT,IER,H)
C
C     DESCRIPTION OF PARAMETERS
C        FUNCT  - USER-WRITTEN SUBROUTINE CONCERNING THE FUNCTION TO
C                 BE MINIMIZED. IT MUST BE OF THE FORM
C                 SUBROUTINE FUNCT(N,ARG,VAL,GRAD)
C                 AND MUST SERVE THE FOLLOWING PURPOSE
C                 FOR EACH N-DIMENSIONAL ARGUMENT VECTOR ARG,
C                 FUNCTION VALUE AND GRADIENT VECTOR MUST BE COMPUTED
C                 AND, ON RETURN, STORED IN VAL AND GRAD RESPECTIVELY
C          N      - NUMBER OF VARIABLES
C          X      - VECTOR OF DIMENSION N CONTAINING THE INITIAL
C                 ARGUMENT WHERE THE ITERATION STARTS. ON RETURN,
C                 X HOLDS THE ARGUMENT CORRESPONDING TO THE
C                 COMPUTED MINIMUM FUNCTION VALUE
C          F      - SINGLE VARIABLE CONTAINING THE MINIMUM FUNCTION
C                 VALUE ON RETURN, I.E. F=F(X).
C          G      - VECTOR OF DIMENSION N CONTAINING THE GRADIENT
C                 VECTOR CORRESPONDING TO THE MINIMUM ON RETURN,
C                 I.E. G=G(X).
C          EST    - IS AN ESTIMATE OF THE MINIMUM FUNCTION VALUE.
C          EPS    - TESTVALUE REPRESENTING THE EXPECTED ABSOLUTE ERROR.
C                 A REASONABLE CHOICE IS 10**(-6), I.E.
C                 SOMEWHAT GREATER THAN 10**(-D), WHERE D IS THE
C                 NUMBER OF SIGNFICANT DIGITS IN FLOATING POINT
C                 REPRESENTATION.
C        LIMIT  - MAXIMUM NUMBER OF ITERATIONS.
C     IER       - ERROR PARAMETER
C                 IER = 0 MEANS CONVERGENCE WAS OBTAINED
C                 IER = 1 MEANS NO CONVERGENCE IN LIMIT ITERATIONS
C                 IER =-1 MEANS ERRORS IN GRADIENT CALCULATION
C                 IER = 2 MEANS LINEAR SEARCH TECHNIQUE INDICATES
C                 IT IS LIKELY THAT THERE EXISTS NO MINIMUM.
C          H      - WORKING STORAGE OF DIMENSION N*(N+7)/2.
C
C     REMARKS
C         I) THE SUBROUNTINE NAME REPLACING THE DUMMY ARGUMENT FUNCT
C            MUST BE DECLARED AS EXTERNAL IN THE CALLING PROGRAM.
C        II) IER IS SET TO 2 IF , STEPPING IN ONE OF THE COMPUTED
C            DIRECTIONS, THE FUNCTION WILL NEVER INCREASE WITHIN
C            A TOLERABLE RANGE OF ARGUMENT.
C            IER = 2 MAY OCCUR ALSO IF THE INTERVAL WHERE F
C            INCREASES IS SMALL AND THE INITIAL ARGUMENT WAS
C            RELATIVELY FAR AWAY FROM THE MINIMUM SUCH THAT THE
C            MINIMUM WAS OVERLEAPED. THIS IS DUE TO THE SEARCH
C            TECHNIQUE WHICH DOUBLES THE STEPSIZE UNTIL A POINT
C            IS FOUND WHERE THE FUNCTION INCREASES.
```

```
C
C     SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED
C        FUNCT
C
C     METHOD
C     THE METHOD IS DESCRIBED IN THE FOLLOWING ARTICLE
C        R. FLETCHER AND M.J.D. POWELL, A RAPID DESCENT METHOD FOR
C        MINIMIZATION,
C        COMPUTER JOURNAL VOL.6, ISS. 2, 1963, PP.163-168.
C
C
C
      COMMON / LINKS / ERR1, X0(60), DIFF(60)
C
C        DIMENSIONED DUMMY VARIABLES
      DIMENSION H(1),X(1),G(1)
C
C        COMPUTE FUNCTION VALUE AND GRADIENT VECTOR FOR INITIAL ARGUMENT
      CALL FUNCT(N,X,F,G)
C
C        RESET ITERATION COUNTER AND GENERATE IDENTITY MATRIX
      IER=0
      KOUNT=0
      N2=N+N
      N3=N2+N
      N31=N3+1
    1 K=N31
      DO 4 J=1,N
      H(K)=1.
      NJ=N-J
      IF(NJ)5,5,2
    2 DO 3 L=1,NJ
      KL=K+L
    3 H(KL)=0.
    4 K=KL+1
C
C        START ITERATION LOOP
    5 KOUNT=KOUNT +1
C
C        SAVE FUNCTION VALUE, ARGUMENT VECTOR AND GRADIENT VECTOR
      OLDF=F
      DO 9 J=1,N
      K=N+J
      H(K)=G(J)
      K=K+N
      H(K)=X(J)
C
C        DETERMINE DIRECTION VECTOR H
      K=J+N3
      T=0.
      DO 8 L=1,N
      T=T-G(L)*H(K)
      IF(L-J)6,7,7
    6 K=K+N-L
      GO TO 8
```

```
    7 K=K+1
    8 CONTINUE
    9 H(J)=T
C
C        CHECK WHETHER FUNCTION WILL DECREASE STEPPING ALONG H.
      DY=0.
      HNRM=0.
      GNRM=0.
C
C        CALCULATE DIRECTIONAL DERIVATIVE AND TESTVALUES FOR DIRECTION
C        VECTOR H AND GRADIENT VECTOR G.
      DO 10 J=1,N
      HNRM=HNRM+ABS(H(J))
      GNRM=GNRM+ABS(G(J))
   10 DY=DY+H(J)*G(J)
C
C        REPEAT SEARCH IN DIRECTION OF STEEPEST DESCENT IF DIRECTIONAL
C        DERIVATIVE APPEARS TO BE POSITIVE OR ZERO.
      IF(DY)11,51,51
C
C        REPEAT SEARCH IN DIRECTION OF STEEPEST DESCENT IF DIRECTION
C        VECTOR H IS SMALL COMPARED TO GRADIENT VECTOR G.
   11 IF(HNRM/GNRM-EPS)51,51,12
C
C        SEARCH MINIMUM ALONG DIRECTION H
C
C        SEARCH ALONG H FOR POSITIVE DIRECTIONAL DERIVATIVE
   12 FY=F
      ALFA=2.*(EST-F)/DY
      AMBDA=1.
C
C        USE ESTIMATE FOR STEPSIZE ONLY IF IT IS POSITIVE AND LESS THAN
C        1. OTHERWISE TAKE 1. AS STEPSIZE
      IF(ALFA)15,15,13
   13 IF(ALFA-AMBDA)14,15,15
   14 AMBDA=ALFA
   15 ALFA=0.
C
C        SAVE FUNCTION AND DERIVATIVE VALUES FOR OLD ARGUMENT
   16 FX=FY
      DX=DY
C
C        STEP ARGUMENT ALONG H
      DO 17 I=1,N
   17 X(I)=X(I)+AMBDA*H(I)
C
C        COMPUTE FUNCTION VALUE AND GRADIENT FOR NEW ARGUMENT
      CALL FUNCT(N,X,F,G)
      FY=F
C
C        COMPUTE DIRECTIONAL DERIVATIVE DY FOR NEW ARGUMENT. TERMINATE
C        SEARCH, IF DY IS POSITIVE, IF DY IS ZERO THE MINIMUM IS FOUND
      DY=0.
      DO 18 I=1,N
   18 DY=DY+G(I)*H(I)
```

```
              IF(DY)19,36,22
C
C                TERMINATE SEARCH ALSO IF THE FUNCTION VALUE INDICATES THAT
C                A MINIMUM HAS BEEN PASSED
           19 IF(FY-FX)20,22,22
C
C                REPEAT SEARCH AND DOUBLE STEPSIZE FOR FURTHER SEARCHES
           20 AMBDA=AMBDA+ALFA
              ALFA=AMBDA
C                END OF SEARCH LOOP
C
C                TERMINATE IF THE CHANGE IN ARGUMENT GETS VERY LARGE
              IF(HNRM*AMBDA-1.E10)16,16,21
C
C                LINEAR SEARCH TECHNIQUE INDICATES THAT NO MINIMUM EXISTS
           21 IER=2
              RETURN
C
C                INTERPOLATE CUBICALLY IN THE INTERVAL DEFINED BY THE SEARCH
C                ABOVE AND COMPUTE THE ARGUMENT X FOR WHICH THE INTERPOLATION
C                POLYNOMIAL IS MINIMIZED
           22 T=0.
           23 IF(AMBDA)24,36,24
           24 Z=3.*(FX-FY)/AMBDA+DX+DY
              ALFA=AMAX1(ABS(Z),ABS(DX),ABS(DY))
              DALFA=Z/ALFA
              DALFA=DALFA*DALFA-DX/ALFA*DY/ALFA
              IF(DALFA)51,25,25
           25 W=ALFA*SQRT(DALFA)
              ALFA=(DY+W-Z)*AMBDA/(DY+2.*W-DX)
              XM = 0.
              DO 255 I = 1, N
              XN = ( T - ALFA ) * H(I) / X(I)
              IF ( ABS( XN ) .GT. XM ) XM = ABS( XN )
255           CONTINUE
              IF ( XM .LT. EPS ) GO TO 36
              DO 26 I=1,N
           26 X(I)=X(I)+(T-ALFA)*H(I)
C
C                TERMINATE, IF THE VALUE OF THE ACTUAL FUNCTION AT X IS LESS
C                THAN THE FUNCTION VALUES AT THE INTERVAL ENDS. OTHERWISE REDUCE
C                THE INTERVAL BY CHOOSING ONE END-POINT EQUAL TO X AND REPEAT
C                THE INTERPOLATION. WHICH END-POINT IS CHOOSEN DEPENDS ON THE
C                VALUE OF THE FUNCTION AND ITS GRADIENT AT X
C
              CALL FUNCT(N,X,F,G)
              IF(F-FX)27,27,28
           27 IF(F-FY)36,36,28
           28 DALFA=0.
              DO 29 I=1,N
           29 DALFA=DALFA+G(I)*H(I)
              IF(DALFA)30,33,33
           30 IF(F-FX)32,31,33
           31 IF(DX-DALFA)32,36,32
           32 FX=F
```

```
          DX=DALFA
          T=ALFA
          AMBDA=ALFA
          GO TO 23
   33 IF(FY-F)35,34,35
   34 IF(DY-DALFA)35,36,35
   35 FY=F
          DY=DALFA
          AMBDA=AMBDA-ALFA
          GO TO 22
C
C        COMPUTE DIFFERENCE VECTORS OF ARGUMENT AND GRADIENT FROM
C        TWO CONSECUTIVE ITERATIONS
   36 DO 37 J=1,N
          K=N+J
          H(K)=G(J)-H(K)
          K=N+K
   37 H(K)=X(J)-H(K)
C
C        TERMINATE, IF FUNCTION HAS NOT DECREASED DURING LAST ITERATION
          IF(OLDF-F+EPS)51,38,38
C
C        TEST LENGTH OF ARGUMENT DIFFERENCE VECTOR AND DIRECTION VECTOR
C        IF AT LEAST N ITERATIONS HAVE BEEN EXECUTED. TERMINATE, IF
C        BOTH ARE LESS THAN EPS
   38 IER=0
          IF(KOUNT-N)42,39,39
   39 T=0.
          Z=0.
          DO 40 J=1,N
          K=N+J
          W=H(K)
          K=K+N
          T=T+ABS(H(K))
   40 Z=Z+W*H(K)
          IF(HNRM-EPS)41,41,42
   41 IF(T-EPS)56,56,42
C
C        TERMINATE, IF NUMBER OF ITERATIONS WOULD EXCEED LIMIT
   42 IF(KOUNT-LIMIT)43,50,50
C
C        PREPARE UPDATING OF MATRIX H
   43 ALFA=0.
          DO 47 J=1,N
          K=J+N3
          W=0.
          DO 46 L=1,N
          KL=N+L
          W=W+H(KL)*H(K)
          IF(L-J)44,45,45
   44 K=K+N-L
          GO TO 46
   45 K=K+1
   46 CONTINUE
          K=N+J
```

```
      ALFA=ALFA+W*H(K)
   47 H(J)=W
C
C        REPEAT SEARCH IN DIRECTION OF STEEPEST DESCENT IF RESULTS
C        ARE NOT SATISFACTORY
      IF(Z*ALFA)48,1,48
C
C        UPDATE MATRIX H
   48 K=N31
      DO 49 L=1,N
      KL=N2+L
      DO 49 J=L,N
      NJ=N2+J
      H(K)=H(K)+H(KL)*H(NJ)/Z-H(L)*H(J)/ALFA
   49 K=K+1
      GO TO 5
C        END OF ITERATION LOOP
C
C        NO CONVERGENCE AFTER LIMIT ITERATIONS
   50 IER=1
      IF ( GNRM - EPS ) 55, 55, 56
C
C        RESTORE OLD VALUES OF FUNCTION AND ARGUMENTS
   51 DO 52 J=1,N
      K=N2+J
   52 X(J)=H(K)
      CALL FUNCT(N,X,F,G)
C
C        REPEAT SEARCH IN DIRECTION OF STEEPEST DESCENT IF DERIVATIVE
C     FAILS TO BE SUFFICIENTLY SMALL
      IF(GNRM-EPS)55,55,53
C
C        TEST FOR REPEATED FAILURE OF ITERATION
   53 IF(IER)56,54,54
   54 IER=-1
      GO TO 1
   55 IER=0
   56 RETURN
      END
```

D. Subroutine FUNB

When a prospective minimum point reaches the boundary, it is constrained to remain on the boundary. This requires that the (x,y) coordinates be exchanged for an angle coordinate, θ (azimuth of the point concerned), and the derivative with respect to θ replaces the derivatives with respect to x and y. This is done by FUNB which calls FUNCT and modifies the derivatives from it so that they are directed tangentially to the boundary.

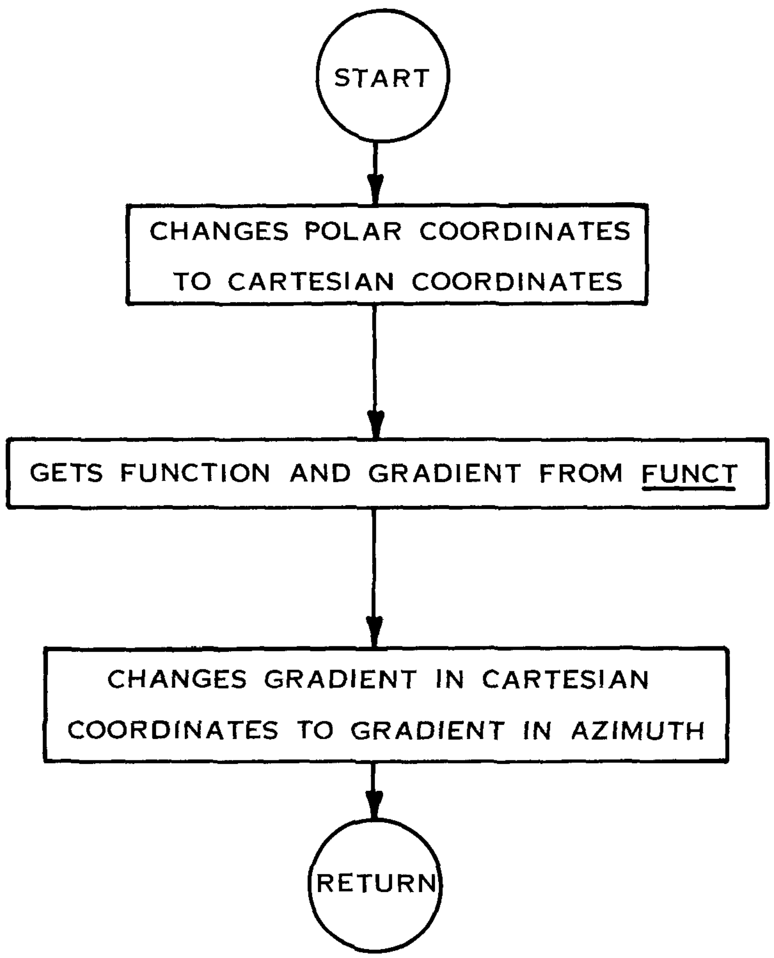


FIGURE V-3

ABBREVIATED FLOW CHART FOR FUNB

```
*DECK FUNB
      SUBROUTINE FUNB( I, ANG, F, G )
C     TO CONVERT FUNCT( N, X, F, G ) INTO A ONE DIMENSIONAL
C     FUNCTION ON THE CIRCULAR BOUNDARY
      COMMON / BLK1 / NSTN, NBDY, NTR, XS(30), YS(30), ITR(50,3)
     , , XC, YC, R, IBDY(20)
      COMMON / Q / IC, IQ
      DIMENSION X(2), G(2), G2(2)
      X(1) = XC + R * COS( ANG )
      X(2) = YC + R * SIN(ANG )
      CALL FUNCT( 2, X, F, G2 )
      G(1) = -G2(1) * ( X(2) - YC ) + G2(2) * ( X(1) - XC )
      RETURN
      END
```

E. Subroutine FUNCT

Subroutine FUNCT(N,X,F,G) is used by FMFP to get the function value and its partial derivatives at the point X(I), I=1, N, where N is the number of independent variables X(I) that are the arguments of the function, F. The function value is scalar and is returned in F. Its partial derivatives with respect to X(I) are returned in the gradient vector G(I), I=1, N.

The function value and its partial derivatives are evaluated using the subroutine CORFUN (q.v.) from the proper values/functions of the correlation coefficient at discrete points and from the field of pollution concentration variance W(I,J), I,J=1, 9, which is available in INT2D. via COMMON/BLK2/.

The computation process is carried out on the following basis. Let a_{ij} be the elements of the correlation coefficient matrix for station locations i, j; let g_i be the correlation coefficient of pollution concentration between stations i and the point (x_o, y_o), the starting point for the minimization procedure (or some point during the search for the minimum); let σ^2 be the variance of pollution concentration at (x_o, y_o). The least square error of estimate of pollution concentration at (x_o, y_o) from linear regression already located stations i, at (x_i, y_i), is given by

$$F = \overline{\varepsilon^2} = \sigma^2 (1 - \sum_i \sum_j g_i a^{ij} g_j) \qquad (1)$$

which is the function value required. Its partial derivatives are given by

$$G(1) = \partial\overline{\varepsilon^2}/\partial x = 2\sigma(\partial\sigma/\partial x)(1 - \sum_i \sum_j g_i a^{ij} g_j)$$

$$- 2\sigma^2 \sum_i \sum_j g_i a^{ij} (\partial g_j/\partial x) \qquad (2)$$

$$G(2) = \partial\overline{\varepsilon^2}/\partial y = 2\sigma(\partial\sigma/\partial y)(1 - \sum_i \sum_j g_i a^{ij} g_j)$$

$$- 2\sigma^2 \sum_i \sum_j g_i a^{ij} (\partial g_j/\partial y) \qquad (3)$$

where a^{ij} are the elements of the inverse of the matrix $\{a_{ij}\}$.

The correlation coefficient function obtained from CORFUN is such that if $c(x,y;\xi,\eta)$ represents the correlation between pollution concentrations at (x,y) and (ξ,η), then the limit for $\xi \to x$, $\eta \to y$ is not 1, but a value somewhat less than 1 (a jump discontinuity at $\xi=x$, $\eta=y$). This means that the correlation function is represented in the form

$$C(x,y;\xi,\eta) = C^*(x,y;\xi,\eta) + A(x,y)\delta(x,y;\xi,\eta) \qquad (4)$$

where $C^*(x,y;\xi,\eta)$ is the continuous part of the correlation coefficient function, $A(x,y)$ is the amount of the jump discontinuity, $\delta(x,y;\xi,\eta)$ is the two-dimensional Dirac function which is zero if $x \neq \xi$, $y \neq \eta$ and is 1 at $x = \xi$, $y = \eta$. Thus, the limit for $\xi \to x, \eta \to y$ of $C^*(x,y;\xi,\eta)$ exists and is $C^*(x,y;x,y) = C_o^*$. In the limit sense $C(x,y;\xi,\eta) \to C_o^*$ but the actual value of $C(x,y;x,y)$ is $C_o^* + A(x,y)$. This means that in (1) the terms of g_i have this jump discontinuity so that $g_i = g(x_i,y_i;x_o,y_o) \to C_o^*(x_i,y_i)$ in the limit sense, but should actually take the value $C_o^*(x_i y_i) + A(x_i y_i)$ when $x_o = x_i$, $y_o = y_i$. Thus, $F(x_o,y_o)$ in the limit sense does not approach zero when (x_o,y_o) approaches the location of an already located pollution concentration observation station.

To force the function to zero in such circumstances the following modified form of the error of estimate (1) was used. Let F_o^2 be the value computed from

$$F_o^2 = 1 - \sum_i \sum_j g_i a^{ij} g_j .$$

Then

$$F = \sigma^2 (F_o - A/F_1)^2$$

will be zero when the point (x_o, y_o) approaches an already located station.

In lines 17-38 the matrix $\{a_{ij}\}$ is set up and inverted. The terms g_i are computed in lines 39-45, σ^2 is obtained on line 48, the error variance on line 55 and the value of F on line 62. If only the function value is required, IC = 1, and the rest of the subroutine is skipped. When derivatives are required, the derivative of the variance of pollution concentration is computed in lines 69-76, the correlation coefficient derivatives are then obtained (77-83) and the remaining terms of the gradient vector are added (78-90).

When diagnostics have been returned by any of the subroutines, their sense is printed (lines 91-106). Since the coordinates of already located stations remain fixed, provision is made in IQ to hold the values of matrix $\{a^{ij}\}$ (IQ = 2) during a given minimization, but to renew them (IQ = 1) when a new minimization is started.

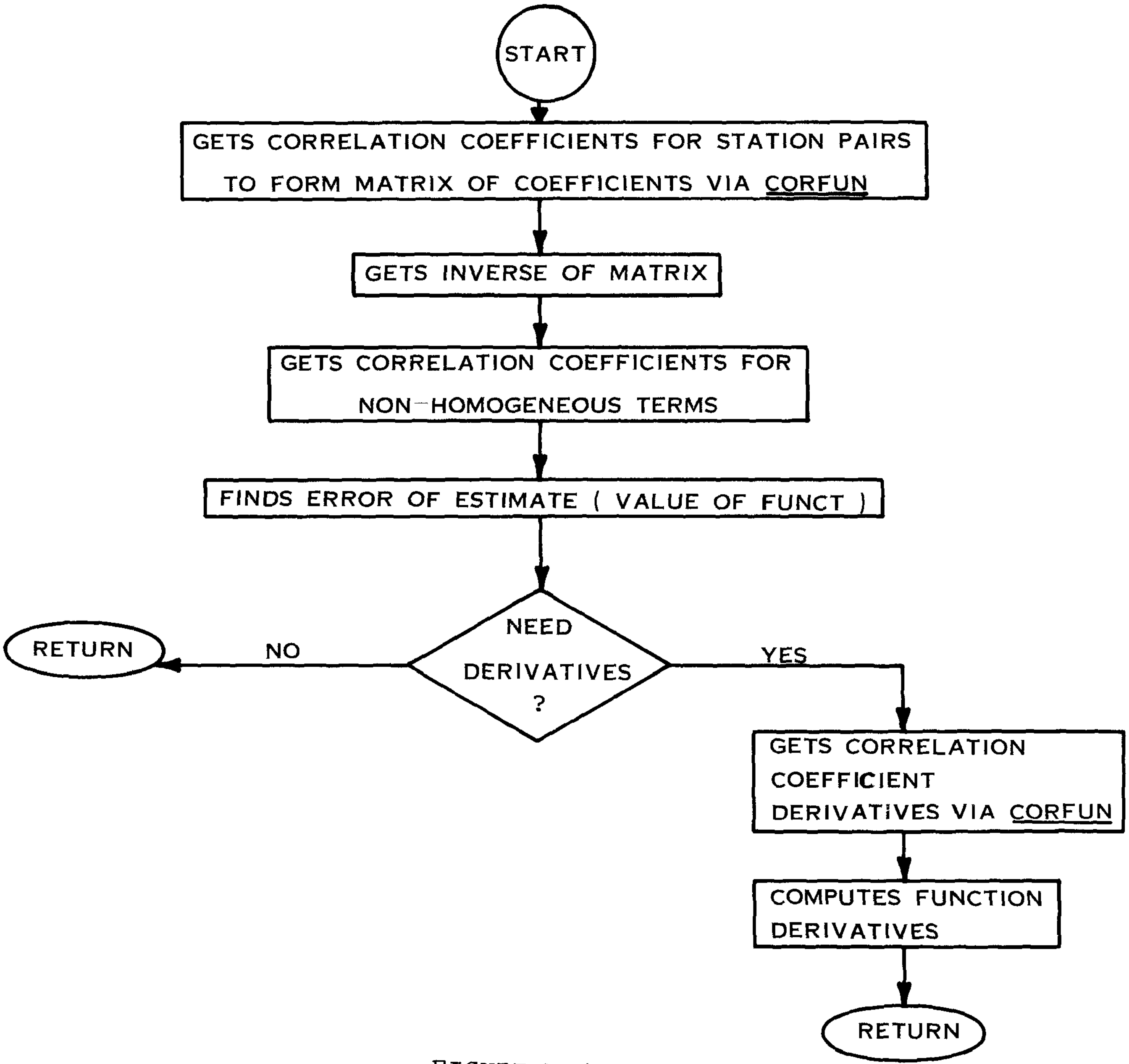


FIGURE V-4

ABBREVIATED FLOW CHART OF FUNCT.

```
*DECK FUNCT
      SUBROUTINE FUNCT( N, X, F, G )
C     TO COMPUTE THE FUNCTION TO BE MINIMIZED
C     N = NO. OF VARIABLES IN ARGUMENT; N = 2
C     X(I), I = 1, N, ARGUMENT
C     F = FUNCTION VALUE ( SCALAR )
C     G(I), I = 1, N, GRADIENT OF FUNCTION
C     IC, CONTROL; 1 = GET F ONLY, 2 = GET F AND G(I)
C     IQ = CONTROL, 1 = DO WHOLE THING, 2 = USE OLD APAM, AA, AINV.
C     SINCE BASIC STATION LOCATIONS ARE NOT CHANGED.
C     NO PENALTY FUNCTION INVOLVED
C
      COMMON / BLK1 / NSTN, NBDY, NTR, XS(30), YS(30), ITR(50,3)
     , , XC, YC, R, IBDY(20)
      COMMON / Q / IC, IQ
      DIMENSION G(2), AINV(30,30), GG(30), DGDX(30), DGDY(30), AA(30,30)
      DIMENSION X(1)
      X0 = X(1)
      Y0 = X(2)
C     GETS CORRELATION COEFFICIENT ARRAYS
      GO TO ( 30, 32 ), IQ
30    DO 6 I = 1, NSTN
      XM = XS(I)
      YM = YS(I)
      DO 4 J = 1, I
      X1 = XS(J)
      Y1 = YS(J)
      CALL CORFUN( XM, YM, X1, Y1, COR, DCDX, DCDY, ID )
      IF ( ID .NE. 0 ) GO TO 102
      AA(I,J) = COR
      IF ( I .EQ. J ) AA(I,J) = 1.
4     CONTINUE
6     CONTINUE
C     FILLS OUT CORCOEF MATRIX AND GETS INVERSE, ETC.
      DO 8 I = 1, NSTN
      J1 = I + 1
      DO 8 J = J1, NSTN
8     AA(I,J) = AA(J,I)
      CALL MATINV( AA, AINV, NSTN, 40 )
C     GET NONHOMOGENEOUS TERMS
32    DO 36 I = 1, NSTN
      XM = XS(I)
      YM = YS(I)
      CALL CORFUN( XM, YM, X0, Y0, COR, DCDX, DCDY, ID )
      IF ( ID .NE. 0 ) GO TO 102
36    GG(I) = COR
      CALL INT2D( X0, Y0, VAL, ID )
      IF ( ID .NE. 0 ) GO TO 104
      VSQ = VAL * VAL
C     COMPUTES ERROR OF ESTIMATE AND ASSIGNS NEGATIVE SIGN SINCE
C     SUBROUTINE FMFP GOES FOR MIN AND WE WANT MAX
      SUM = 0.
      DO 10 I = 1, NSTN
      DO 10 J = 1, NSTN
10    SUM = SUM + GG(I) * AINV(I,J) * GG(J)
```

```
              ERR = 1. - SUM
              IF ( ERR .LT. 0. ) GO TO 106
      11      CONTINUE
              CALL CORFUN1( X0, Y0, X0, Y0, COR, DCDX, DCDY, ID )
              A = 1. - COR
              F0 = SQRT( ERR )
              IF ( F0 .EQ. 0. ) F = 0.
              IF ( F0 .NE. 0. ) F = -VSQ * ( F0 - A / F0 )**2
              PRINT 500, ERR, COR, A, F0, F
      500     FORMAT (5X,*F0 = *,E13.5, * COR = *,E13.5,* A = *,E13.5, * SQRT(F0
             *) = *,E13.5,* F1 - F0 = *,E13.5)
              IF ( IC .EQ. 1 ) GO TO 200
      C       ESTIMATES DERIVATIVES USING 1 UNIT INCREMENT ON THE
      C       WEIGHT FUNCTION
              X1 = X0 + 1.
              Y1 = Y0 + 1.
              CALL INT2D( X1, Y0, VX, ID )
              IF ( ID .NE. 0 ) GO TO 104
              CALL INT2D( X0, Y1, VY, ID )
              IF ( ID .NE. 0 ) GO TO 104
              G(1) = -2. * VAL * ( VX - VAL ) * ERR
              G(2) = -2. * VAL * ( VY - VAL ) * ERR
              DO 22 I = 1, NSTN
              XM = XS(I)
              YM = YS(I)
              CALL CORFUN( XM, YM, X0, Y0, COR, DCDX, DCDY, ID )
              IF ( ID .NE. 0 ) GO TO 102
              DGDX(I) = DCDX
      22      DGDY(I) = DCDY
              SUM1 = SUM2 = 0.
              DO 24 I = 1, NSTN
              DO 24 J = 1, NSTN
              SUM1 = SUM1 + GG(I) * AINV(I,J) * DGDX(J)
      24      SUM2 = SUM2 + GG(I) * AINV(I,J) * DGDY(J)
              G(1) = G(1) + 2. * VSQ * SUM1
              G(2) = G(2) + 2. * VSQ * SUM2
      C       THESE GRADIENT COMPNENTS HAVE SIGN REVERSED TO CORRESPOND TO F.
      200     RETURN
      C       DIAGNOSTICS PRINT OUTS
      102     PRINT 1002, ID
      1002    FORMAT (* COORDINATE *,I3,* OUT OF RANGE IN CORFUN FROM SUBROUTINE
             * FUNCT*)
              GO TO 300
      104     PRINT 1004, ID
      1004    FORMAT (* COORDINATE *, I3, * OUT OF RANGE IN INT2D FROM SUBROUTIN
             *E FUNCT*)
              GO TO 300
      106     CONTINUE
              PRINT 1006, ERR, X0, Y0
      1006    FORMAT (* INTERPOLATION ERR *,E10.3,* FROM SUBROUTINE FUNCT AT X =
             * *,F10.5,*  Y = *,F10.3)
              ERR = 0.
              GO TO 11
      300     CALL EXIT
              END
```

F. Subroutine CORFUN

Subroutine CORFUN is used to determine the correlation coefficient relating pollution concentration at two arbitrary points neither of which need be a point of the 9 x 9 grid. The technique used is that of a quadratic interpolation procedure based on the description of the correlation function in terms of its proper functions and values. If the proper values are λ_i, i=1, ..., k, $(\lambda_1 > \lambda_2 > \ldots > \lambda_k)$ and the proper functions corresponding thereto are $\phi_i(x,y)$, then the correlation function may be expressed in the form

$$K(x,y;\xi,\eta) = \sum_{i=1}^{k} \lambda_i \phi_i(x,y) \phi_i(\xi,\eta) \qquad (1)$$

where (x,y) and (ξ,η) are the points between which pollution concentration is being correlated. The value of k is that of the statistically significant proper values/functions. The proper functions are known only at the points of a 9 x 9 square grid of points (10 km separation). The values of $\phi(x,y)$ and $\phi(\xi,\eta)$ are found by means of a simple bivariate quadratic interpolation procedure.

Had the correlation coefficient function itself been used, $K(x,y;\xi,\eta)$ a four variable interpolation method would be required and additional constraints imposed to preserve the basic character of the correlation coefficient function (such as having a value not exceeding +1 and a "horizontal" tangent plane at any point $x = \xi$, $y = \eta$). The use of the proper functions greatly simplifies the interpolation procedure and guarantees that the characteristics of the correlation function are preserved.

The interpolation algorithm is standard and is given by M. Abramowitz and I.A. Stegun, Handbook of Mathematical Functions, U.S. Government Printing Office, Washington, D.C., June 1964, p. 882, formula 25.2.67. This requires six points for a rectangular grid in the arrangement of Figure V-5. The formula is

$$f(x_0+ph,y_0+qk) = q(q-1)f_{0,-1}/2+p(p-1)f_{-1,0}/2+pqf_{1,1}$$

$$+(1+pq-p^2-q^2)f_{0,0}+p(p-2q+1)f_{1,0}/2+q(q-2p+1)f_{0,1}/2$$

which has been rearranged into the form

$$f(x_0+ph,y_0+qk) = f_{0,0}+(1/2)[p(f_{+1,0}-f_{-1,0})+q(f_{0,+1}-f_{0,-1}) +$$

$$+p^2(f_{+1,0}+f_{-1,0}-2f_{0,0})+q^2(f_{0,+1}+f_{0,-1}-2f_{0,0})]$$

$$+pq(f_{0,0}+f_{+1,+1}-f_{+1,0}-f_{0,+1}).$$

In the above, (h,k) are the lengths of the sides of the rectangular grid, and $f_{i,j}$ are function values at the grid points with units indicated by the subscript and scale factors h for index i, k for index j. The values of p and q are determined from $p=(x-x_0)/h$, $q=(y-y_0)/k$.

The arrangement of points in Figure V-5 provides for values of p and q in the ranges $0\leq p\leq 0.5$, $0\leq q\leq 0.5$. For the point (p,q) lying in the other quadrants of the rectangle with corners (0,0), (+1,0), (+1,+1), (0,+1) are shown in Figure V-6.

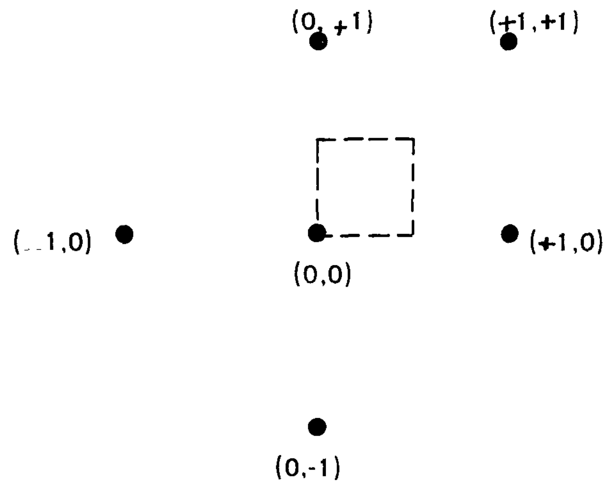


FIGURE V-5

THE ARRANGEMENT OF DATA POINTS FOR 6-POINT QUADRATIC INTERPOLATION. THE AREA OF MOST ACCEPTABLE VALUES OF P AND Q ($0<P<0.5$) IS THE DOTTED SQUARE WITH (0,0) AS ITS LOWER LEFT CORNER.

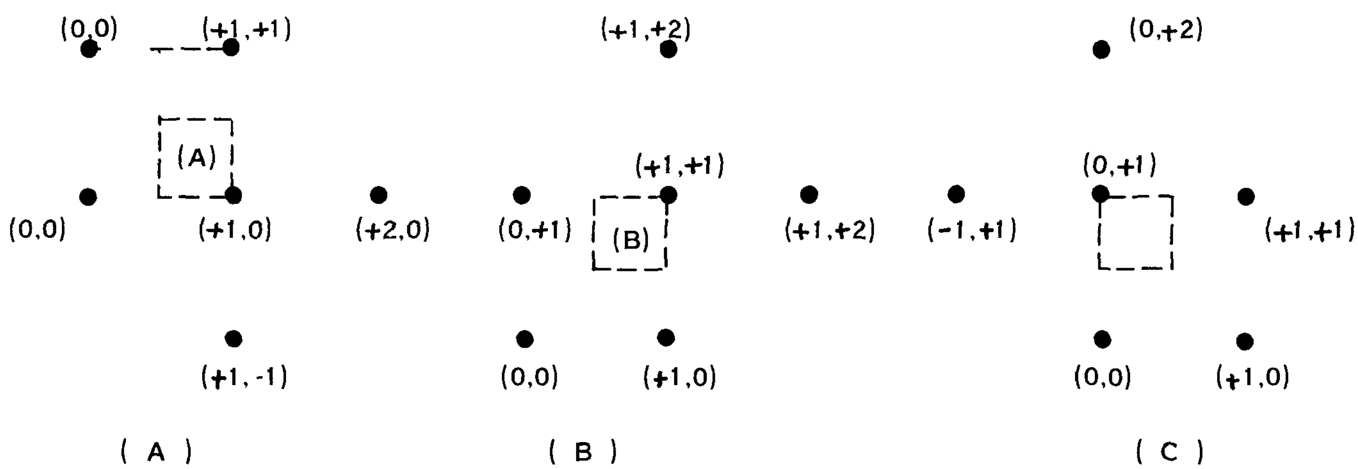


FIGURE V-6

EQUIVALENT ARRANGEMENTS OF DATA POINTS TO ACCOMMODATE OTHER VALUES OF P, Q LYING CLOSE TO THE POINT (XO,YO)

The partial derivatives of the correlation coefficient function are obtained by differentiating (1),

$$\partial K(x,y;\xi,\eta)/\partial\xi = \sum_{i=1}^{k} \lambda_i \phi_i(x,y)\,[\partial\phi_i(\xi,\eta)/\partial\xi]$$

$$\partial K(x,y;\xi,\eta)/\partial\eta = \sum_{i=1}^{k} \lambda_i \phi_i(x,y)\,[\partial\phi_i(\xi,\eta)/\partial\xi]$$

so that the result requires the differentiating of the proper functions in terms of which the correlation function is expressed. Since quadratic interpolation was used, the formula for the interpolated derivatives comes immediately from differentiation of the interpolation formula. Thus

$$\partial f(x_0+ph,y_0+qk)/\partial p = 0.5(f_{+1,0}-f_{-1,0})+p(f_{+1,0}+f_{-1,0}-2f_{0,0})$$

$$+q(f_{0,0}+f_{+1,+1}-f_{+1,0}-f_{0,+1}),$$

$$\partial f(x_0+ph,y_0+qk)/\partial q = 0.5(f_{0,+1}-f_{0,-1})+p(f_{0,0}+f_{+1,+1}-f_{+1,0}-f_{0,+1})$$

$$+q(f_{0,+1}+f_{0,-1}-2f_{0,0}).$$

The input points at which the correlation function values are required in CORFUN(XM,YM,X1,Y1,COR,DCDX,DCDY,ID) are inserted as XM, YM and X1,Y1 corresponding to (x,y) and (ξ,η) above. The function value is output as COR and the derivatives as DCDX and DCDY. The diagnostic ID has the value 0 if all is well. It takes on values +1 if XM is out of range, +2 if YM is out of range, +3 if X1 is out of range, and +4 if Y1 is out of range. The proper values and proper functions are obtained from COMMON/BLK2/ as ALAM(K) and PHI(K,I,J) and have been tabulated at the points with coordinates XD(I), YD(J). COMMON/Q/ contains a control parameter IC such that if IC=1 the calculation of derivatives is omitted and if IC=2 it is included.

```
*DECK CORFUN
      SUBROUTINE CORFUN( XM, YM, X1, Y1, COR, DCDX, DCDY, ID )
C     GETS CORRELATION FUNCTIONS AND DERIVATIVES WO X1, Y1 FROM TABLES
C     OF PROPER FUNCTIONS / VALUES.
C     ID = 0 FOR COORDINATES IN RANGE, ID .NE. 0, OUT OF RANGE
      DIMENSION PH1(15), PHM(15), DPH1DX(15), DPH1DY(15)
      COMMON / BLK2 / XD(9), YD(9), PH(15,9,9), ALAM(15), W(9,9)
      COMMON / Q / IC, IQ
C     PH(L2,I,J) = L2TH PROPER FUNCTION AT X(I), Y(J)
C     ALAM(L2) = L2TH PROPER VALUE
C
      DATA L2 / 15 /
      ID = 0
      IF ( XM .EQ. X1 .AND. YM .EQ. Y1 ) 4, 6
4     COR = 1. $ DCDX = 0. $ DCDY = 0.
      GO TO 100
      ENTRY CORFUN1
6     XA = XM $ YA = YM $ K = 1
C     LOCATE LOWER LEFT CORNER OF SQUARE CONTAINING XA, YA
8     IF ( XA .LT. XD(1) ) GO TO 11
      DO 10 I = 2, 9
      IF ( XA .LE. XD(I) ) GO TO 12
10    CONTINUE
11    ID = 1 + 2 * ( K - 1 )
      GO TO 100
12    IF ( YA .LT. YD(1) ) GO TO 14
      DO 13 J = 2, 9
      IF ( YA .LE. YD(J) ) GO TO 16
13    CONTINUE
14    ID = 2 + 2 * ( K - 1 )
      GO TO 100
16    IZ = I - 1 $ JZ = J - 1
C     INTERPOLATION ON PROPER FUNCTIONS
      P0= .1 * ( XA - XD(IZ) )
      Q0= .1 * ( YA - YD(JZ) )
      DO 34 L = 1, L2
      IF ( P0.LE. .5 ) 18, 20
18    IF ( Q0.LE. .5 ) 22, 24
20    IF ( Q0.LE. .5 ) 26, 28
22    ICASE = 1 $ P = P0 $ Q = Q0
      A = PH(L,IZ,JZ) $ B = PH(L,IZ+1,JZ) $ C = PH(L,IZ-1,JZ)
      D = PH(L,IZ,JZ+1) $ E = PH(L,IZ,JZ-1) $ F = PH(L,IZ+1,JZ+1)
      GO TO 30
24    ICASE = 3 $ P = P0 $ Q = 1. - Q0
      A = PH(L,IZ,JZ+1) $ B = PH(L,IZ+1,JZ+1) $ C = PH(L,IZ-1,JZ+1)
      D = PH(L,IZ,JZ) $ E = PH(L,IZ,JZ+2) $ F = PH(L,IZ+1,JZ)
      GO TO 30
26    ICASE = 2 $ P = 1. - P0 $ Q = Q0
      A = PH(L,IZ+1,JZ) $ B = PH(L,IZ,JZ) $ C = PH(L,IZ+2,JZ)
      D = PH(L,IZ+1,JZ+1) $ E = PH(L,IZ+1,JZ-1) $ F = PH(L,IZ,JZ+1)
      GO TO 30
28    ICASE = 4 $ P = 1. - P0$ Q = 1. - Q0
      A = PH(L,IZ+1,JZ+1) $ B = PH(L,IZ,JZ+1) $ C = PH(L,IZ+2,JZ+1)
      D = PH(L,IZ+1,JZ) $ E = PH(L,IZ+1,JZ+2) $ F = PH(L,IZ,JZ)
30    IF ( K .NE. 1 ) PHM(L) = PH1(L)
```

```
              PH1(L) = A + .5 * ( P * ( B - C ) + Q * ( D - E ) + P * F * ( B +
             + C - 2. * A ) + Q * Q * ( D + E - 2. * A ))+ P * Q * ( A +F-B-D )
              IF ( K .EQ. 1 ) GO TO 34
              GO TO ( 34, 32 ), IC
        32    SGNP = 1. $ SGNQ = 1.
              IF ( ICASE .EQ. 3 .OR. ICASE .EQ. 4 ) SGNQ = -1.
              IF ( ICASE .EQ. 2 .OR. ICASE .EQ. 4 ) SGNP = -1.
              DPH1DX(L) = SGNP * ( .5 * ( B - C ) + P * ( B + C - 2. * A ) +
             + Q * ( A + F - B - D ) )
              DPH1DY(L) = SGNQ * ( .5 * ( D - E ) + Q * ( D + E - 2. * A ) +
             + P * ( A + F - B - D ) )
        34    CONTINUE
              GO TO ( 36, 38 ), K
        36    K = 2
              XA = X1 $ YA = Y1
              GO TO 8
        38    COR = 0. $ DCDX = 0. $ DCDY = 0.
              DO 42 L = 1, L2
              COR = COR + PHM(L) * ALAM(L) * PH1(L)
              GO TO ( 42, 40 ), IC
        40    DCDX = DCDX + PHM(L) * ALAM(L) * DPH1DX(L)
              DCDY = DCDY + PHM(L) * ALAM(L) * DPH1DY(L)
        42    CONTINUE
        100   RETURN
              END
```

G. Subroutine INT2D

This is a two-dimensional interpolation subroutine using the standard formula

$$f = f_{0,0}+p(f_{1,0}-f_{0,0})+q(f_{0,1}-f_{0,0})+pq(f_{0,0}-f_{0,1}-f_{1,0}+f_{1,1})$$

Four points $f_{0,0}$, $f_{0,1}$, $f_{1,0}$, $f_{1,1}$ at the corners of a rectangle are used. The values p and q are given by $p = (x_a-x_1)/(x_2-x_1)$, $q = (y_a-y_1)/(y_2-y_1)$ where (x_a,y_a) is the point at which the value of f is required and (x_1,y_1), (x_2,y_1), (x_1,y_2), (x_2,y_2) are the coordinates at the corners of the rectangle. The surface fitted to the data is a hyperbolic paraboloid.

In the form INT2D(XA,YA,VAL,IDIOG) the coordinates (x_a,y_a) are input as XA, YA, the result is output in VAL. IDIOG=0 for successful interpolation, =1 for XA outside the table range, =2 for YA outside the table range. In this case the table, W(I,J), I,J=1,9, appears in the COMMON statement. The quantity PH(15,9,9) of the COMMON statement is not used in this subroutine.

```
*DECK INT2D
      SUBROUTINE INT2D(XA,YA,VAL,IDIOG)
      COMMON / BLK2 /  X(9),  Y(9), PH(15,9,9), ALAM(15), W(9,9)
C     INTERPOLATES VALUE OF W AT XA,YA, FROM VALUES GIVEN AT X,Y
C     N=9, IDIOG = DIAGNOSTIC, 0=OK, 1=X OUT OF RANGE, 2= Y OUT OF RANGE
      IDIOG=0
      DO 10 I=1,9
      IF(XA.LT.X(I)) GO TO 12
   10 CONTINUE
      IDIOG=1
      GO TO 18
   12 DO 14 J=1,9
      IF(YA.LT.Y(J)) GO TO 16
   14 CONTINUE
      IDIOG=2
      GO TO 18
   16 I2=I  $  J2=J
      I1=I-1  $  J1=J-1
      P=(XA-X(I1))/(X(I2)-X(I1))
      Q=(YA-Y(J1))/(Y(J2)-Y(J1))
      W1=W(I1,J1)
      W2=W(I2,J1)
      W3=W(I1,J2)
      W4=W(I2,J2)
      VAL=W1+P*(W2-W1)+Q*(W3-W1)+P*Q*(W1-W2-W3+W4)
   18 RETURN
      END
```

H. Subroutine MATINV

This is a standard subroutine for matrix inversion. In MATINV(A,B,N,M) the matrix to be inverted is contained in A and N gives the number of rows/columns of A. The inverse of A is returned in B. Both A and B are listed in full form.

```
*DECK MATINV
      SUBROUTINE MATINV(A,B,N,M)                                        MATI   0
      DIMENSION A(30,30),B(30,30),IPIVOT(40),JPIVOT(40),C(40)
C                                                                       MATI  20
C     INITIALIZATION                                                    MATI  30
C                                                                       MATI  40
      DO 15 I=1,N                                                       MATI  50
      DO 10 J=1,N                                                       MATI  60
   10 B(I,J)=A(I,J)                                                     MATI  70
      IPIVOT(I)=0                                                       MATI  80
   15 JPIVOT(I)=0                                                       MATI  90
      DO 115 II=1,N                                                     MATI 100
C                                                                       MATI 110
C     SEARCH FOR PIVOT ELEMENT                                          MATI 120
C                                                                       MATI 130
      PIVOT=0.                                                          MATI 140
      DO 60 I=1,N                                                       MATI 150
      IF(IPIVOT(I).NE.0) GO TO 60                                       MATI 160
      DO 50 J=1,N                                                       MATI 170
      IF(JPIVOT(J).NE.0) GO TO 50                                       MATI 180
      IF(ABS(PIVOT).GT.ABS(B(I,J))) GO TO 50                            MATI 190
      PIVOT = B(I,J)                                                    MATI 200
      IROW=I                                                            MATI 210
      JCOL=J                                                            MATI 220
   50 CONTINUE                                                          MATI 230
   60 CONTINUE                                                          MATI 240
      IPIVOT(IROW)=JCOL                                                 MATI 250
      JPIVOT(JCOL)=IROW                                                 MATI 260
C                                                                       MATI 270
C     REPLACE PIVOT COLUMN WITH ROW MULTIPLIERS                         MATI 280
C                                                                       MATI 290
      DO 70 I=1,N                                                       MATI 300
      X=-B(I,JCOL)                                                      MATI 310
      IF(IROW.NE.I) GO TO 70                                            MATI 320
   65 X=1.                                                              MATI 330
   70 B(I,JCOL)=X/PIVOT                                                 MATI 340
C                                                                       MATI 350
C     REDUCE NON PIVOT COLUMNS                                          MATI 360
C                                                                       MATI 370
      DO  90 I=1,N                                                      MATI 380
   90 C(I) = B(IROW,I)                                                  MATI 390
      DO 115 J=1,N                                                      MATI 400
      IF(JCOL.EQ.J) GO TO 115                                           MATI 410
   95 DO 110 I=1,N                                                      MATI 420
      IF(IROW.EQ.I) GO TO 105                                           MATI 430
  100 B(I,J) = B(I,J)+B(I,JCOL)*C(J)                                    MATI 440
      GO TO 110                                                         MATI 450
  105 B(I,J)=B(I,J)/PIVOT                                               MATI 460
  110 CONTINUE                                                          MATI 470
  115 CONTINUE                                                          MATI 480
C                                                                       MATI 490
C     INTERCHANGE ROWS AND COLUMNS                                      MATI 500
C                                                                       MATI 510
      DO 130 I=1,N                                                      MATI 520
      DO 120 K=1,N                                                      MATI 530
```

```
      L = IPIVOT(K)                                                     MATI 540
  120 C(K) = B(I,L)                                                     MATI 550
      DO 130 K=1,N                                                      MATI 560
  130 B(I,K) = C(K)                                                     MATI 570
      DO 150 I=1,N                                                      MATI 580
      DO 140 K=1,N                                                      MATI 590
      L = JPIVOT(K)                                                     MATI 600
  140 C(K) = B(L,I)                                                     MATI 610
      DO 150 K=1,N                                                      MATI 620
  150 B(K,I) = C(K)                                                     MATI 630
      RETURN                                                            MATI 640
      END                                                               MATI 650
```

I. Subroutine ADDPT(XO,YO)

This subroutine adds the station location with coordinates XO,YO to the station list and rearranges the list so that it is in canonical form, i.e., the triangle assignments include the new point and vertices of all triangles are in counterclockwise order; if it is a new boundary point it is listed as such and the number of boundary points incremented and the counterclockwise ordering of the boundary points is maintained.

The logic is somewhat involved and many exceptional cases occur. The subroutine print-out contains many explanatory comments on what is taking place. These should be sufficient to unravel the situation. Extensive use is made of subroutines AR2, TRITST, PORDR, and TRIFIX.

One of the basic ideas is to the effect that if XO,YO is the point P and if I1, I2 and neighboring boundary points listed in counterclockwise order, then the "area" of the triangle formed by the ordered points P, I1, I2 will be "positive" (i.e., P, I1, I2 are in counterclockwise order around a triangle) for all boundary point pairs I1, I2 (neighboring) if P lies inside the boundary, and will be "negative" for at least some pair I1, I2 if P is outside the boundary, and will be "zero" if P lies on I1, I2 (interior or exterior). If the point is clearly interior to the boundary of the points already located, it is required then to find the triangle of already located points within which it lies (or on a side of which it lies). One then subdivides this triangle. If the new point lies outside the boundary of already located points, then additional triangles are to be formed. The many exceptional cases require careful handling since experience indicates that they occur with distressing frequency.

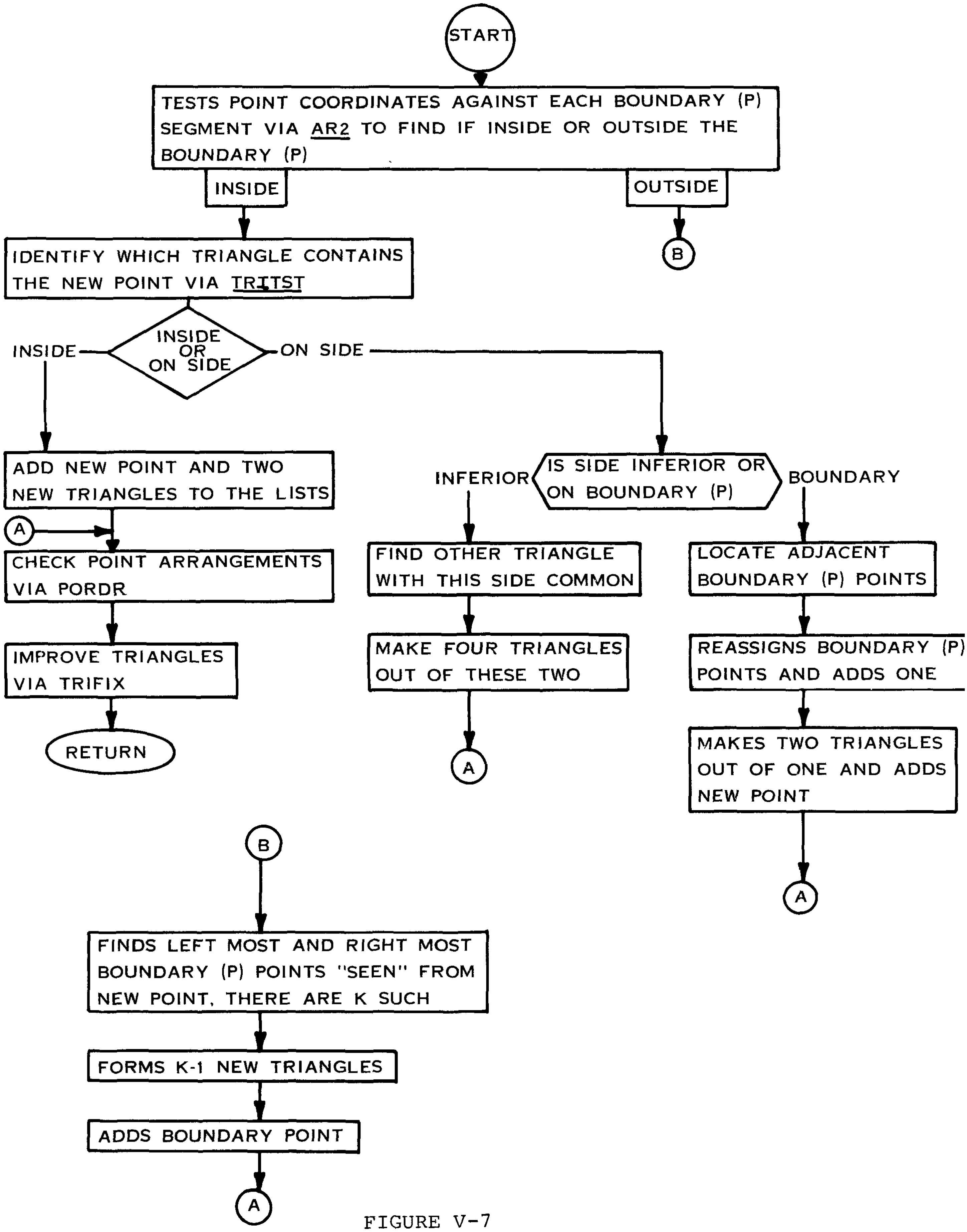


FIGURE V-7

ABBREVIATED FLOW CHART FOR ADDPT

```
*DECK ADDPT
      SUBROUTINE ADDPT( X0, Y0 )
C     ADDS POINT TO AN ARRAY IN STANDARD FORM SO THAT THE STANDARD FORM
C     IS PRESERVED AND CHECKS TRIANGLE ARRANGEMENTS
C     CHECKS TRIANGLE ARRANGEMENT SO IT IS ,,GOOD,,.
      COMMON / BLK1 / NSTN, NBDY, NTR, XS(30), YS(30), ITR(50,3)
     , , XC, YC, R, IBDY(20)
      INTEGER P
      DIMENSION ISIG(20), P(2), IH(20)
C     PUTS ( X0, Y0 ) IN XS(NSTN+1), YS(NSTN+1)
      N1 = NSTN + 1
      XS(N1) = X0
      YS(N1) = Y0
C     CHECKS WHETHER ( X0, Y0 ) IS INSIDE OR OUTSIDE THE BOUNDARY
      DO 10 I = 1, NBDY
      I1 = I + 1 $ IF ( I .EQ. NBDY ) I1 = 1
      CALL AR2( IBDY(I), IBDY(I1), N1, A, ID )
      IF ( ID .LE. 0 ) GO TO 54
10    CONTINUE
C     ALL VALUES OF ID ARE + SO IS INSIDE.  FIND WHICH TRIANGLE
      DO 12 I = 1, NTR
      J1 = ITR(I,1) $ J2 = ITR(I,2) $ J3 = ITR(I,3)
      CALL TRITST( X0, Y0, J1, J2, J3, ID, A, P, Q )
      IF ( ID .GT. 0 ) GO TO 14
12    CONTINUE
C     POINT IS NOT AN ABSOLUTE INTERIOR POINT.  CHECK AGAIN TO SEE IF
C     IS ON AN INTERIOR BOUNDARY
      DO 13 I = 1, NTR
      J1 = ITR(I,1) $ J2 = ITR(I,2) $ J3 = ITR(I,3)
      CALL TRITST( X0, Y0, J1, J2, J3, ID, A, P, Q )
      IF ( ID .GT. -7 ) GO TO 14
13    CONTINUE
      PRINT 1065
1065  FORMAT (* SEARCH FOR TRIANGLE DID NOT CONFIRM THAT NEW POINT WAS A
     *N INTERIOR POINT*)
      GO TO 126
14    IT = I
      IF ( ID .LE. 0 ) GO TO 20
C     HAVE FOUND TRIANGLE WITH ( X0, Y0 ) AS INTERIOR POINT
C     ADDS TWO MORE TRIANGLES
      IA = ITR(IT,1) $ IB = ITR(IT,2) $ IC = ITR(IT,3)
      ITR(IT,3) = N1
      I4 = NTR + 1
      ITR(I4,1) = IB $ ITR(I4,2) = IC $ ITR(I4,3) = N1
      I4 = I4 + 1
      ITR(I4,1) = IC $ ITR(I4,2) = IA $ ITR(I4,3) = N1
C     JOB DONE.  TWO TRIANGLES, ONE POINT, NO BDRY PTS ADDED
      GO TO 120
C      NOW TAKE CARE OF EXCEPTIONAL CASES
20    IF ( ID .LT. 0 ) GO TO 22
C     HAVE A DEGENERATE TRIANGLE
      PRINT 1070, I, J1, J2, J3
1070  FORMAT (* TRIANGLE *,I5,* VERTICES*,3I5,* IS A DEGENERATE*)
      GO TO 126
C     THE POINT ( X0, Y0 ) IS ON A SIDE OF THE TRIANGLE
```

```
C      FIND THE OTHER TRIANGLE WITH THIS SIDE AND MAKE 4 OF 2.
22     ID = -ID
       GO TO ( 24, 28, 32, 36, 40, 42 ), ID
24     IA = J1 $ IB = J3 $ IC = J2 $ GO TO 44
28     IA = J3 $ IB = J1 $ IC = J2 $ GO TO 44
32     IA = J2 $ IB = J1 $ IC = J3 $ GO TO 44
36     IA = J1 $ IB = J2 $ IC = J3 $ GO TO 44
40     IA = J3 $ IB = J2 $ IC = J1 $ GO TO 44
42     IA = J2 $ IB = J3 $ IC = J1
C      POINT ( X0, Y0 ) LIES ON SIDE IA, IB.  FIND ANOTHER TRIANGLE WITH
C      THIS SIDE
44     DO 48 I = 1, NTR
       IF ( I .EQ. IT ) GO TO 48
       DO 46 J = 1, 3
       J1 = J + 1 $ IF ( J .EQ. 3 ) J1 = 1
       IF ( IA .EQ. ITR(I,J) .AND. IB .EQ. ITR(I,J1) ) GO TO 50
       IF ( IB .EQ. ITR(I,J) .AND. IA .EQ. ITR(I,J1) ) GO TO 50
46     CONTINUE
48     CONTINUE
C      THE TRIANGLE MUST HAVE SIDE IA, IB ON THE BOUNDARY
       GO TO 85
50     J2 = J + 2 $ IF ( J2 .GT. 3 ) J2 = J2 - 3
       IE = ITR(I,J2) $ ITA = I
C      NOW MAKES FOUR OUT OF TWO
       ITR(IT,1) = IC $ ITR(IT,2) = IA $ ITR(IT,3) = N1
       ITR(ITA,1) = IA $ ITR(ITA,2) = IE $ ITR(ITA,3) = N1
       N2 = NTR + 1
       ITR(N2,1) = IB $ ITR(N2,2) = IC $ ITR(N2,3) = N1
       N2 = N2 + 1
       ITR(N2,1) = IE $ ITR(N2,2) = IB $ ITR(N2,3) = N1
C      JOB DONE, TWO TRIANGLES, ONE POINT, NO BDRY PTS ADDED
       GO TO 120
C      POINT IS OUTSIDE OR ON THE BOUNDARY OR ON
C      AN EXTENSION OF A BOUNDARY SIDE
C      CHECKS SIGN FOR ORDER OF POINTS I1, I2, N1
C      THIS + IF CC, - IF CW
54     DO 56 I = 1, NBDY
       I1 = I + 1 $ IF ( I .EQ. NBDY ) I1 = 1
56     CALL AR2( IBDY(I), IBDY(I1), N1, A, ISIG(I) )
C      +1 IF N1 DOES NOT ,,SEE,, SIDE I1, I2, -1 IF IT DOES,
C      0 IF N1 LIES ON I1, I2 OR I1, I2 EXTENDED
C      CHECK FOR EXCEPTIONAL CASES
       DO 58 I = 1, NBDY
       IF ( ISIG(I) .NE. 0 ) GO TO 58
       IM = I - 1 $ IF ( I .EQ. 1 ) IM = NBDY
       IP = I + 1 $ IF ( I .EQ. NBDY ) IP = 1
       IF ( ( ISIG(IM) .EQ. 1 ) .AND. ( ISIG(IP) .EQ. 1 ) ) GO TO 90
       GO TO 60
58     CONTINUE
       GO TO 62
60     ISIG(I) = 1
C      CHECKS OUT LEFT AND RIGHT POINTS ,,SEEN,, FROM NEW POINT
62     K = 0
       DO 66 I = 1, NBDY
       I1 = I + 1 $ IF ( I .EQ. NBDY ) I1 = 1
```

```
          IF ( ISIG(I) .EQ. ISIG(I1) ) 66, 64
    64    K = K + 1
          P(K) = I1
    66    CONTINUE
          IF ( ISIG(P(1) ) .EQ. -1 ) 68, 70
    68    IL = P(1)
          IR = P(2)
          GO TO 72
    70    IL = P(2)
          IR = P(1)
    C      IL = FIRST POINT TO LEFT ,,SEEN,, FROM NEW POINT
    C     IR = LAST POINT TO RIGHT ,,SEEN,, FROM NEW POINT
    C     K = NUMBER OF POINTS ,,SEEN,, FROM NEW POINT
    72    K = IR - IL + 1
          IF ( K .LT. 0 ) K = NBDY + K
          IF ( K .LE. 1 ) 74, 76
    74    I1 = IBDY(IL) $ I2 = IBDY(IR)
          PRINT 1075, I1, I2, X0, Y0
    1075  FORMAT (* LEFT AND RIGHT BOUNDARY POINTS*/2I5/*ARE SEEN FROM*/
         ,2F10.5)
          GO TO 126
    C     THERE ARE K - 1 NEW TRIANGLES
    76    I2 = K- 1 $ N2 = NTR
          DO 78 I = 1, I2
          N2 = N2 + 1
          J1 = IL + I - 1 $ IF ( J1 .GT. NBDY ) J1 = J1 - NBDY
          J2 = J1 + 1 $ IF ( J2 .GT. NBDY ) J2 = J2 - NBDY
          ITR(N2,1) = IBDY(J2)
          ITR(N2,2) = IBDY(J1)
    78    ITR(N2,3) = N1
    C     REASSIGNS BOUNDARY POINTS
    C     PUTS OLD BOUNDARY POINTS IN HOLD
    80    DO 82 I = 1, NBDY
    82    IH(I) = IBDY(I)
    C     ASSIGNS I = 1 TO NEW POINT
          IBDY(1) = N1
          I2 = NBDY + 3 - K
          DO 84 I = 2, I2
          I1 = IR + I - 2 $ IF ( I1 .GT. NBDY ) I1 = I1 - NBDY
    84    IBDY(I) = IH(I1)
          GO TO 122
    C     NEW POINT ON INTERIOR OF BOUNDARY SEGMENT
    C     CASES WHERE SEGMENT NOT SPECIFIED
    85    DO 86 I = 1, NBDY
          I1 = I + 1 $ IF ( I .EQ. NBDY ) I1 = 1
          IF ( IA .EQ. IBDY(I) .AND. IB .EQ. IBDY(I1) ) GO TO 96
          IF ( IB .EQ. IBDY(I) .AND. IA .EQ. IBDY(I1) ) GO TO 88
    86    CONTINUE
          PRINT 1080, X0, Y0
    1080  FORMAT (* FAILS TO LOCATE BOUNDARY SEGMENT CONTAINING POINT*/
         ,2F10.5)
          GO TO 126
    C
    88    IA = IBDY(I) $ IB = IBDY(I1) $ GO TO 96
    C     LOOKS FOR TRIANGLE AND SUBDIVIDES
```

SUBROUTINE ADDPT

```
90    K2 = IBDY(IP)
      I1 = IP - 1 $ IF ( I1 .EQ. 0 ) I1 = 1
      K1 = IBDY(I1)
      DO 92 I = 1, NTR
      DO 92 J = 1, 3
      J1 = J + 1 $ IF ( J .EQ. 3 ) J1 = 1
      IF ( K1 .EQ. ITR(I,J) .AND. K2 .EQ. ITR(I,J1) ) GO TO 94
      IF ( K2 .EQ. ITR(I,J) .AND. K1 .EQ. ITR(I,J1) ) GO TO 94
92    CONTINUE
      PRINT 1085, K1, K2
1085  FORMAT(* FAILS TO LOCATE TRIANGLE WITH SEGMENT*/2I5/* ON EXTERIOR
     *BOUNDARY*)
      GO TO 126
94    J2 = J + 2 $ IF ( J2 .GT. 3 ) J2 = J2 - 3
      IT = I
      IA = ITR(IT,J) $ IB = ITR(IT,J1) $ IC = ITR(IT,J2)
96    ITR(IT,1) = IC $ ITR(IT,2) = IA $ ITR(IT,3) = N1
      IT = NTR + 1
      ITR(IT,1) = IB $ ITR(IT,2) = IC $ ITR(IT,3) = N1
      K = 2
      GO TO 80
120   NSTN = NSTN + 1
      NTR = NTR + 2
      GO TO 124
122   NSTN = NSTN + 1
      NTR = NTR + K - 1
      NBDY = NBDY + 3 - K
124   CALL PORDR
      CALL TRIFIX
200   RETURN
126   PRINT 1090
1090  FORMAT (* ABOVE DIAGNOSTICS FROM SUBROUTINE ADDPT*)
      RETURN
      END
```

J. Subroutine TRIFIX

This subroutine readjusts the triangle assignments to maintain a network of non-overlapping triangles with observation points at their vertices to prevent the occurrence of triangles of unnecessarily small area. It finds triangle pairs with a common side. When such a pair is located, they are combined into a quadrilateral. This quadrilateral may be divided into two ways. The triangle subdivision is prefered that has the shorter of the two diagonals of the quadrilateral as the common side (Figure V-8).

The quadrilateral formed by two triangles with a common side may be re-entrant. In this case the second diagonal lies "outside" the quadrilateral and the subdivision is accepted as is (line 36), see Figure V-9. The case is identified by the fact that when the outside diagonal is used, one of the triangle areas resulting is larger than the sum of the areas of the original triangle pairs.

When the re-entrant quadrilateral test fails but still the area of the quadrilateral obtained by the second subdivision exceeds that obtained by the first subdivision by more than 0.01% a diagnostic is printed. This test is required to account for the effect of roundoff errors in the computer.

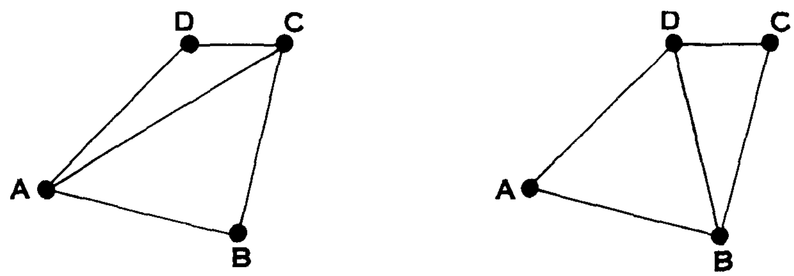


FIGURE V-8

THE TWO POSSIBLE SUBDIVISIONS OF QUADRILATERAL ABCD INTO TWO TRIANGLES. ON THE LEFT, THE SUBDIVISION ABC, ACD RESULTS IN TRIANGLES WITH A LARGER COMMON SIDE AC WHILE THE SUBDIVISION ABD,BCD ON THE RIGHT RESULTS IN A SMALLER COMMON SIDE BD AND IS THEREFORE PREFERRED.

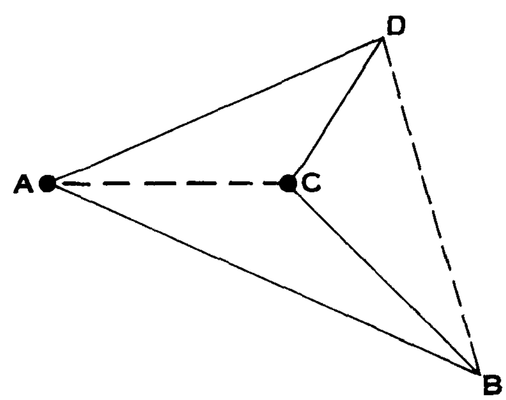


FIGURE V-9

THE REENTRANT QUADRILATERAL CASE. THE ORIGINAL TRIANGLES ARE ABC AND ACD. THE SECOND DIAGONAL BD LIES " OUTSIDE " THE QUADRILATERAL AND IS REJECTED.

```
*DECK TRIFIX
      SUBROUTINE TRIFIX
C     CHECKS THE TRIANGLE SUBDIVISION AND TRYS TO IMPROVE THE
C     TRIANGLE ARRANGEMENTS.  FINDS TRIANGLE PAIRS WITH COMMON SIDE AND
C     TESTS THE DIAGONALS OF THE QUADRILATERAL TO SEE IF MORE NEARLY
C     EQUAL AREA SUBDIVISION IS POSSIBLE
      COMMON / BLK1 / NSTN, NBDY, NTR, XS(30), YS(30), ITR(50,3)
     , , XC, YC, R, IBDY(20)
      DIMENSION X1(3), X2(3), Y1(3), Y2(3)
      I2 = NTR - 1
      DO 20 I = 1, I2
      K1 = I + 1
      DO 20 K = K1, NTR
      DO 16 J = 1, 3
      J1 = J $ J2 = J + 1 $ IF ( J .EQ. 3 ) J2 = 1
      JA = ITR(I,J1) $ JB = ITR(I,J2)
      DO 16 L = 1, 3
      L1 = L $ L2 = L + 1 $ IF ( L .EQ. 3 ) L2 = 1
      LA = ITR(K,L1) $ LB = ITR(K,L2)
      IF ( ( JA .EQ. LA ) .AND. ( JB .EQ. LB ) ) 2, 4
2     ISW = 1 $ GO TO 8
4     IF ( ( JA .EQ. LB ) .AND. ( JB .EQ. LA ) ) 6, 16
6     ISW = 2
8     J3 = J2 + 1 $ IF ( J3 .GT. 3 ) J3 = 1
      L3 = L2 + 1 $ IF ( L3 .GT. 3 ) L3 = 1
      JA = ITR(I,J1) $ JB = ITR(I,J2) $ JC = ITR(I,J3)
      LC = ITR(K,L3)
      GO TO ( 11, 10 ), ISW
10    LA = ITR(K,L2) $ LB = ITR(K,L1)
11    CALL AR2( JA, JB, JC, A1, ID1 )
      CALL AR2( LA, LB, LC, A2, ID2 )
      CALL AR2( JB, JC, LC, A3, ID3 )
      CALL AR2( JC, JA, LC, A4, ID4 )
      AT = A1 + A2
C     TESTS IF QUADRILATERAL REENTRANT.  IF IS CANNOT
C     OTHERWISE SIMPLY DIVIDE
      IF ( ( A3 .GE. AT ) .OR. ( A4 .GE. AT ) ) GO TO 20
      BT = A3 + A4
      IF ( ABS( BT - AT ) .LT. .0001 * AT )  GO TO 14
      DO 12 M = 1, 3
      X1(M) = XS(ITR(I,M)) $ Y1(M) = YS(ITR(I,M))
      X2(M) = XS(ITR(K,M)) $ Y2(M) = YS(ITR(K,M))
12    CONTINUE
      PRINT 1002, I, ( ITR(I,M), X1(M), Y1(M), M = 1, 3 ) ,
     , K, ( ITR(K,M), X2(M), Y2(M), M = 1, 3 )
1002  FORMAT (* THESE TRIANGLES CANNOT BE */2(I5,3(I5,2F10.5,5X)/))
      PRINT 1001, JA,JB,JC,A1,LA,LB,LC,A2,JB,JC,LC,A3,JC,JA,LC,A4,AT,BT
1001  FORMAT (3I3,E12.4,3I3,E12.4,3I3,E12.4,3I3,3E12.4)
      GO TO 20
14    DSQ1 = ( XS(JA) - XS(JB) )**2 + ( YS(JA) - YS(JB) )**2
      DSQ2 = ( XS(JC) - XS(LC) )**2 + ( YS(JC) - YS(LC) )**2
      IF ( DSQ1 .LE. DSQ2 ) GO TO 20
C     SUBDIVISION IMPROVED IF DIAGONALS EXCHANGED
      ITR(I,1) = JC $ ITR(I,2) = LC $ ITR(I,3) = LB
      ITR(K,1) = JC $ ITR(K,2) = JA $ ITR(K,3) = LC
```

SUBROUTINE TRIFIX

```
16    CONTINUE
20    CONTINUE
      RETURN
      END
```

K. Subroutine PORDR

This subroutine checks the arrangement of the vertex points for counterclockwise ordering and rearranges the ordering if a clockwise ordering is found.

```
*DECK POROR
      SUBROUTINE POROR
C     CHECKS THE ARRANGEMENT OF POINTS AROUND THE TRIANGLES TO
C     ASSURE CC ORDERING
      COMMON / BLK1 / NSTN, NBDY, NTR, XS(30), YS(30), ITR(50,3)
      DO 4 I = 1, NTR
      J1 = ITR(I,1) $ J2 = ITR(I,2) $ J3 = ITR(I,3)
      CALL AR2( J1, J2, J3, A, IOIOG )
      IF ( IOIOG .EQ. 1 ) GO TO 4
      ITR(I,1) = J2 $ ITR(I,2) = J1
4     CONTINUE
      RETURN
      END
```

L. Subroutine AR2

This subroutine computes twice the area of the triangle with vertices at the points with indices I,J,K and outputs this as A with a diagnostic IDIOG which takes the value of +1 if I,J,K in counterclockwise order about the triangle, -1 if in clockwise order, and 0 if A=0.

```
*DECK AR2
      SUBROUTINE AR2( I, J, K, A, IDIOG )
C     COMPUTES TWICE THE AREA OF THE TRIANGLE WITH STATIONS I, J, K AS
C     VERTICES.  RETURNS ABSOLUTE VALUE AT A AND SIGN IN IDIOG AS +1 OR
C     -1.  IDIOG = +1 MEANS I, J, K IN CC ORDER.
      COMMON / BLK1 / NSTN, NBDY, NTR, XS(30), YS(30), ITR(50,3)
     , , XC, YC, R, IBDY(20)
      X1 = XS(I) $ X2 = XS(J) $ X3 = XS(K)
      Y1 = YS(I) $ Y2 = YS(J) $ Y3 = YS(K)
      A = ( X2 - X1 ) * ( Y3 - Y1 ) - ( X3 - X1 ) * ( Y2 - Y1 )
      IDIOG = SIGN( 1., A )
      IF ( A .EQ. 0. ) IDIOG = 0.
      A = ABS( A )
      RETURN
      END
```

M. Subroutine TRITST

This subroutine determines the location of the point input, P, with coordinates X, Y with respect to a triangle with vertices at the points J1, J2, J3 where these are the index numbers for the coordinates of an already located observation point. The output consists of a diagnostic (IDIOG), and twice the areas of the following triangles: A for J1,J2,J3, P for J1,P,J3, Q for J1,J2,P. The diagnostic indicates as follows:

IDIOG =	2	, J1,J2,J3 in counterclockwise order, (X,Y) inside
	1	, " " " " clockwise " " "
	0	, " " " in straight line
	-1,-2	, P lies on J1,J3
	-3,-4	, " " " J1,J2
	-5,-6	, " " " J2,J3
	-7,-8	, " is outside opposite J2
	-9,-10	, " " " " J3
	-11,-12	, " " " " J1

```
*DECK TRITST
      SUBROUTINE TRITST( X, Y, J1, J2, J3, IDIOG, A, P, Q )
C     TESTS WHERE P(X,Y) LIES WO VERTICES J1, J2, J3.
C     IDIOG = EVEN, VERTICES IN CC ORDER, = ODD FOR CW ORDER
C     IDIOG = 0, VERTICES IN STRAIGHT LINE
C     IDIOG = 1, 2, (X,Y) INSIDE TRIANGLE;  = -1, -2, ON J1, J3:  = -3,
C     -4, ON J1, J2:  = -5, -6, ON J2, J3;  = -7, -8, OUTSIDE OPPOSITE
C     J2; = -9, -10, OUTSIDE OPPOSITE J3: = -11, -12, OUTSIDE OPPOSITE
C     J1; A = 2 * AREA OF J1J2J3, P = 2* AREA J1PJ3, Q = 2 * AREA J1J2P
      COMMON / BLK1 / NSTN, NBDY, NTR, XS(30), YS(30), ITR(50,3)
     , , XC, YC, R, IBDY(20)
      K = 0
      X1 = XS(J1) $ X2 = XS(J2) $ X3 = XS(J3)
      Y1 = YS(J1) $ Y2 = YS(J2) $ Y3 = YS(J3)
      A1 = X2 - X1 $ A2 = X3 - X1 $ A3 = X - X1
      B1 = Y2 - Y1 $ B2 = Y3 - Y1 $ B3 = Y - Y1
      A = A1 * B2 - A2 * B1
      P = A3 * B2 - B3 * A2
      Q = A1 * B3 - B1 * A3
      IF ( A ) 2, 12, 4
2     K = 1
      A = -A $ P = -P $ Q = -Q
4     IF ( P ) 20, 14, 6
6     IF ( Q ) 22, 16, 8
8     IF ( A - P - Q ) 24, 18, 10
10    IDIOG = 2 - K
      GO TO 26
12    IDIOG = 0
      GO TO 26
14    IDIOG = -2 + K
      IF ( Q .LT. 0. ) GO TO 20
      IF ( Q .GT. A ) GO TO 24
      GO TO 26
16    IDIOG = -4 + K
      IF ( P .GT. A ) GO TO 22
      GO TO 26
18    IDIOG = -6 + K
      GO TO 26
20    IDIOG = -8 + K
      GO TO 26
22    IDIOG = -10 + K
      GO TO 26
24    IDIOG = -12 + K
26    RETURN
      END
```

REFERENCES

1. Abramowitz, M. and Stegun, I.A. (1964), Handbook of Mathematical Functions, N.B.S. Applied Mathematics Series No. 55, Superintendent of Documents, U.S.G.P.O., Washington, D.C., xiv + 1046 pp.

2. Albrecht, J. and L. Collatz (1958), Zur numerischen Auswertung mehrdimensionaler Integrale; Zeitschrift für Angewandte Mathematik und Mechanik, Vol. 38, pp. 1-15.

3. Anderson, T.W. (1958), An Introduction to Multivariate Statistical Analyses, John Wiley & Sons, Inc., New York, xii + 374 pp.

4. Bartlett, M.S. (1951), "The effect of standardization on an approximation in factor analysis", Biometrika, Vol. 38, pp. 337-334.

5. Beard, R.E. (1947), Some Notes on Approximate Product-Integration, Institute of Actuaries, London, Vol. 73, pp. 356-416.

6. Bellman, Richard (1960), Introduction to Matrix Analysis, McGraw-Hill Book Co., Inc., New York, xx + 328 pp.

7. Boland, W. Robert and C.S. Duris (1969), Product Type Quadrature Formulas, Mathematical Report 69-9, Department of Mathematics, Drexel Institute of Technology, Philadelphia, PA 19104, 48 pp.

8. Boland, W. Robert and C.S. Duris (1971), Product Type Quadrature Formulas, B.I.T., Vol. 11, pp. 139-158.

9. Boland, W. Robert (1972), The Convergence of Product-type Quadrature Formulas, SIAM J. Numerical Analysis, Vol. 9, pp. 6-13.

10. Brooks, C.E.P., Durst, C.S., Carruthers, N., and Dewar, D. (1950a), "Upper Winds over the World", Geophysical Memoirs, No. 85, Meteorological Office, Air Ministry, MO499e, London, 57 pp.

11. Brooks, C.E.P., Durst, C.S., and Carruthers, N. (1946), "Upper Winds over the World, Part 1, The frequency distribution of winds at a point in the free air", Quarterly Journal of the Royal Meteorological Society, London, Vol. 72, pp. 55-73.

12. Buell, C. Eugene (1972), "Integral Equation Representation for 'Factor Analysis'", Journal of Atmospheric Science, Vol. 28, No. 8, pp. 1502-1505.

13. Ceschino, F. and M. Letin (1972), Généralisation de la méthode de Havie au calcul des intégrales multiples, C.R. Acad. Sc. Paris, Series A, Vol. 275, pp. 505-508

14. Craddock, J.M. and Flintoff, S. (1970), "Eigenvector Representation of Northern Hemisphere Fields", Quarterly Journal of the Royal Meteorological Society, Vol. 96, pp. 124-129.

15. Davis, Philip J. and Philip Rabinowitz (1967), Numerical Integration, Blaisdell Pub. Co., Waltham, Mass, 225 pp.

16. Dixon, Valerie A. (1973), Numerical Quadrature, A Survey of Available Algorithms, National Physics Laboratory, Division of Numerical Analysis and Computing, Oxford, NPL Report NAC36, 28 pp.

17. Eimutis, E.C., and Konicek, M.G. (1972), "Derivations of continuous functions for the lateral and vertical atmospheric dispersion coefficients", Atmospheric Environment, Vol. 6, pp. 859-863.

18. Ewing, G.M. (1941), On Approximate Cubature, Amer. Math. Monthly, Vol. 48, pp. 134-136.

19. Farmer, S.A. (1971), "An Investigation of the Results of Principal Component Analysis of Data Derived from Random Numbers", The Statistician, Vol. 20, pp. 63-74.

20. Gantmacher, F.R. (1960a), The Theory of Matrices, I, Chelsea Publishing Co., New York, x + 374 pp.

21. Gantmacher, F.R. (1960b), The Theory of Matrices, II, Chelsea Publishing Co., New York, ix + 276 pp.

22. Hammer, P.C. and A.W. Wymore (1957), Numerical Evaluation of Multiple Integrals, I., Mathematics of Computation, Vol. 11, pp. 59-67.

23. Hammer, P.C. and A.H. Stroud (1958), Numerical Evaluation of Multiple Integrals, II, Mathematics of Computation, Vol. 12, pp. 272-280.

24. Horst, Paul (1965), Factor Analysis of Data Matrices, Holt, Rinehart and Winston, Inc., New York, xix + 730 pp.

25. Jöreskog, K.G. (1962), "On the statistical treatment of residuals in factor analysis", Psychometrika, Vol. 27, pp. 335-354.

26. Kenney, J.F. and Keeping, E.S. (1951), Mathematics of Statistics, Part Two, D. Van Nostrand Co., Inc., New York, xii + 429 pp.

27. Lawley, D.N. (1956), "Tests of significance for the latent roots of covariance and correlation matrices, Biometrika, Vol. 43, pp. 128-136.

28. Lawley, D.N. and Maxwell, A.E. (1963), Factor Analysis as a Statistical Method, Butterworths, London, vii + 113 pp.

29. Lawley, D.N. and Maxwell, A.E. (1971), Factor Analysis as a Statistical Method, Butterworths, London, viii + 153 pp.

30. Levy, Paul (1965), Processes Stochastique et Movement Brownien, Gauthier-Villars et Co., Paris, vi + 438 pp.

31. MacDuffee, C.C. (1949), Vectors and Matrices, The Mathematical Association of America, Providence, R.I., xi + 203 pp.

32. Mehta, M.L. (1967), Random Matrices and the Statistical Theory of Energy Levels, Academic Press, New York, x + 259 pp.

33. Mises, R. von(1936), Formules de cubature, Revere Mathematique del'Union Interbalkanique, Vol. 1, pp. 2-27.

34. Mises, R. von (1954), Numerische Berechnung mehrdimensionaler Integrale, Zeitschrift für Angewemdte Math. Und Mech., Vol. 34, pp. 201-210.

35. Panchev, S. (1971), Random Functions and Turbulence, Pergamon Press, Oxford, xiii + 444 pp.

36. Perlis, S. (1952), Theory of Matrices, Addison-Wesley Press, Inc., Cambridge, Mass., xiv + 237 pp.

37. Ruff, Ronald E (1973a), Memorandum of January 10 with attachments.

38. Ruff, Ronald E (1973b), Memorandum of 23 January with attachments.

39. Slepian, David (1962), "The one sided barrier problem for Gaussian Noise", The Bell System Technical Journal, Vol. 41, March 1962, pp. 463-501.

40. Stroud, A.H. (1971), Approximate Calculation of Multiple Integrals, Prentice-Hall, Inc., Englewood Cliffs, N.J., 418 pp.

41. Stroud, A.H. (1960a), Quadrature Methods for Functions of More than one Variable, Annals of the New York Academy of Sciences, Vol. 86, pp. 776-791.

42. Stroud, A.H. (1960b), Numerical Integration Formulas of Degree Two, Mathematics of Computation, Vol. 14, pp. 21-26.

43. Spearman, Carl (1904), "General intelligence objectively determined and measured", Amer. J. Psychol., Vol. 15, pp. 201-293.

44. Synge, J.L. (1953), Simple bounding formulas for integrals, Canadian Journal of Mathematics, Vol. 5, pp. 46-52.

45. Thacher, H.C. (1957), Optimum Quadrature Formulas in 8 Dimensions, Mathematical Tables and Other Aids to Computation, Vol. 11, pp. 189-194.

46. Turnbull, H.W. (1960), The Theory of Determinants, Matrices and Invariants, Dover Publications, Inc., New York, xviii + 374 pp. (From the second edition of the original, 1945).

47. Turner, D. Bruce (1970), Workbook of Atmospheric Dispersion Estimates, Air Resources Field Research Office, Environmental Science Services Administration, U.S. Department of Health, Education and Welfare, National Air Pollution Central Administration, Cincinnati, Ohio, vii + 84 pp.

48. Tyler, G.W. (1953), Numerical Integration of Functions of Several Variables, Canadian Journal of Mathematics, Vol. 5, pp. 393-412.

49. Wiener, N. (1949), Extrapolation, Interpolation, and Smoothing of Stationary Time Series, The Technology Press and John Wiley & Sons, New York, ix + 163 pp.

APPENDIX A

Conditions on an Empirical Formula for a Correlation Coefficient

Before adopting the use of the proper function/proper value expansion of the correlation coefficients as described in Chapter III, it was felt that an empirical formula could be developed that would represent the correlation coefficients involved. This effort was unsuccessful. At least a part of the reasons for this lie in the following analysis of the structure of correlation coefficients or covariances.

An empirical formula to represent a correlation coefficient field may not be chosen in a perfectly arbitrary way even though the expression leads to values that are confined to the range (+1,-1) and has a maximum of +1 at zero separation between the points concerned. There are several side conditions that the functions must satisfy. They are stated here in terms of only one independent variable to keep the notation reasonably simple, but may be extended to two or three independent variables with no difficulty. We consider the correlation of a field property, p, at two points, x_1 and x_2, or at x and $x+\xi$. Thus the correlation coefficient concerned will be written as $r(x_2,x_1)$, $r(x+\xi,x)$, or as $r(\xi;x)$ all of which will be considered as equivalent.

The usual conditions that must be satisfied are:

1. The correlation coefficient takes on the value +1 when $x_1=x_2$ or at $\xi=0$; i.e., $r(x_1,x_1)=+1$, $r(x+0,x)=+1$, $r(0;x)=+1$. On the other hand it is not necessary that the correlation coefficient be continuous at this point. It may have a jump discontinuity here. In other words, it may be that the limit, for $\xi\to 0$, may be less than +1, $\lim r(\xi;x)=\alpha$, $\alpha<1$.

2. The value of the correlation coefficient must be within the range (+1,-1).

3. It must satisfy the symmetry condition $r(x_2,x_1)=r(x_1,x_2)$. This may also be expressed as $r(x+\xi,x)=r(x,x+\xi)$ and also as $r(\xi;x)=r(-\xi;x+\xi)$. This condition is occasionally overlooked.

4. It must be a positive definite (or at least a non-negative definite) function. In the case of an homogeneous field, this implies that its Fourier transform be positive (or non-negative).

Elementary statistics texts do not usually discuss these points in detail. The reader is referred to Cramer and Leadbetter (1967), Panchev (1971), and others. Slepian (1962) is particularly interesting though it deals principally with another subject. Levy (1965) is especially good.

In constructing an empirical formula for a correlation coefficient, one is tempted to use an expression in the form $r(\xi;x)=r(\xi;\alpha(x),\beta(x),---)$ where $\alpha(x)$, $\beta(x)$ are parameters which in turn are functions of the location of the point x.

To understand what goes on, consider a simple series expansion of the correlation coefficient function

$$r(\xi,x) = 1+A_2(x)\xi^2+A_3(x)\xi^3+ \dots$$

where P_1 has coordinate x and P_2 is at $x+\xi$. The coefficient $A_2(x)$ is negative. If we reverse the roles of P_1 and P_2, then P_1 will have coordinate $(x+\xi)-\xi$ and P_2 will be $x+\xi$ so that one has $r(\xi,x)=r(-\xi,x+\xi)$ for the equality $r(P_2,P_1)=r(P_1,P_2)$. Then one must have equality of the two series

$$1+A_2(x)\xi^2+A_3(x)\xi^3+\ldots = 1+A_2(x+\xi)\xi^2-A_3(x+\xi)\xi^3+\ldots \quad .$$

If the coefficients also have valid power series expansion about x (or $\xi=0$), we expand coefficients on the right and collect terms in ascending powers of ξ to obtain

$$1+A_2\xi^2+A_3\xi^3+A_4\xi^4 =$$

$$= 1+A_2\xi^2+(A_2'-A_3)\xi^3+(A_2''/2!-A_3'+A_4)\xi^4$$

$$+ (A_2'''/3!-A_3''/2!+A_4'-A_5)\xi^5+\ldots$$

Equating coefficients of like powers of ξ on each side, then

$$A_3 = -A_3 + A_2'$$

$$A_4 = A_4 - A_3' + A_2''/2!$$

$$A_5 = -A_5 + A_4' - A_3''/2! + A_2'''/3!$$

$$A_6 = A_6 - A_5' + A_4''/2! - A_3'''/3! + A_2^{IV}/4!$$

--- ---

It is readily seen that these lead to a system of equations for the coefficients of the odd order terms in terms of the derivatives of the even ordered terms, and that they give no information on the even order terms. If we write

$$A_{2n-1} = a_1 A'_{2n-2} + a_2 A'''_{2n-4} + \ldots + a_n A_2^{(2n-1)}$$

The coefficients a_i are given by the system of equations

$$2a_1 = 1/1!$$

$$a_1/2!+2a_2 = 1/3!$$

$$a_1/4!+a_2/2!+2a_3 = 1/5!$$

$$a_1/6!+a_2/4!+a_3/2!+2a_4 = 1/7!$$

--- ---

$$a_1/(2n-2)!+a_2(2n-4)!+ \ldots +a_{n-1}/2!+a_{2n} = 1/(2n-1)!$$

--- ---

which are easily solved recursively, the first few solutions are $a_1=1/2!$, $a_2=-1/4!$, $a_3=3/6!$, $a_4=-17/8!$, $a_5=155/10!$, etc. The general expression for the solution is not immediately obvious.

A stochastic field of property is said to be homogeneous (in the extended sense) if its statistical parameters are independent of the location of the point P_1 (here, x) with which the points P_2 (here, $x+\xi$) are correlated. This means that the correlation coefficient $r(\xi;x)$ is a function of ξ alone, $r(\xi)$, and (of course) the standard deviations are also constant.

The fact that the correlation coefficient for a non-homogeneous process contains odd order terms in its series expansion does not imply that a function form with odd order terms necessarily represents a non-homogeneous process. For example, nearly all simple empirical expressions for a correlation coefficient are applicable only to an homogeneous process. The very popular expression $r=\exp(-|\xi|/L)$, L = scale parameter, can only apply to an homogeneous process. Thus, if we assume the contrary and let $r(x_2,x_1)=\exp(-|x_2-x_1|/L(x_1))$, then $r(x_1,x_2)=\exp(-|x_1-x_2|/L(x_2))$, and since these correlation coefficients are equal, it follows at once that $L(x_1)=L(x_2)$ which means that L is a constant and hence that the process is homogeneous.

The same kind of analysis may be made for the two point covariance functions. It is not necessary that the covariance function have a maximum at the point $P_1=P_2$. In the case of a one-dimensional function one may use

$$\mathrm{cov}(P_2,P_1) = \mathrm{cov}(\xi;x) = \mathrm{cov}(x)+C_1(x)\xi+C_2(x)\xi^2 + ---$$

and

$$\mathrm{cov}(P_1,P_2) = \mathrm{cov}(-\xi;x+\xi) = C_0(x+\xi)-C_1(x+\xi)\xi+C_2(x+\xi)\xi^2 + ---$$

$$= C_0+C_o'\xi+C_0''\xi^2/2!+C_0'''\xi^3/3!+C_0^{IV}\xi^4/4!+ \ldots$$

$$-\xi(C_1+C_1'\xi+C_1''\xi^2/2!+C_1'''\xi^3/3!+ \ldots$$

$$+\xi^2(C_2+C_2'\xi+C_2''\xi^2/2!+ \ldots)-\xi^3(C_3+C_3' + \ldots) + \ldots$$

Now rearranging in terms of powers of ξ,

$$\mathrm{cov}(P_1,P_2) = C_0+(C_0'-C_1)\xi+(C_0''/2!-C_2')\xi^2+(C_0'''/3!-C_1''/2!+C_2'-C_3)$$

$$+ (C_0^{IV}/4!-C_1'''/3!+C_2''/2!-C_3'+C_4)\xi^4+ \ldots$$

Equating coefficients in the first and last of these expressions since $\mathrm{cov}(P_2,P_1) = \mathrm{cov}(P_1,P_2)$ one obtains the system of relations

$$C_0 = C_o$$

$$C_1 = -C_1+C_0'$$

$$C_2 = C_2-C_1'+C_0''/2!$$

$$C_3 = -C_3+C_2'-C_1''/2!+C_0'''/3!$$

$$C_4 = C_4-C_3'+C_2''/2!-C_1'''/3!+C_0^{IV}/4!$$

which lead to expressions for the odd order terms as

$$2C_1 = C_0'$$

$$2C_3 = C_2'' - C_1''/2! + C_0'''/3!$$

$$2C_5 = C_4' - C_3''/2! + C_3'''/3! - C_1^{IV}/4! + C_0^{V}/5!$$

--- ---

This system of equations is the same as those obtained for the terms in the expansion of the correlation coefficient with the exception that one starts here with C_1 while before one started with A_3 and we have rather general values for C_0 and C_1 while in the previous example $A_1=0$ and for any correlation coefficient $A_0=1$.

The point of these exercises is to emphasize the fact that care must be exercised in selecting an empirical formula to represent a correlation coefficient function. Otherwise one may have an expression that cannot represent a correlation coefficient function. This is particarly the case when the field concerned is not homogeneous.

APPENDIX B

NOTE ON THE RANK OF A COVARIANCE MATRIX

Let A be the covariance matrix and let X be a data matrix in which the element x_{ij} is the departure from the mean of the pollutant concentration at station i on day j. Let there be n stations (i=1, ---, n) and d reporting days (j=1, ---, d). Then the covariance matrix may be written in the form

$$A = XX'/d, \qquad X' = \text{transpose of } X$$

so that the element a_{ij} is the mean sum of products

$$a_{ij} = \left(\sum_{k=1}^{d} x_{ik}x_{jk}\right)/d.$$

and is the covariance of pollutant concentration at stations i and j. The matrix A is nxn while X is nxd and X' is dxn.

It is a well known theorem that the rank of a matrix cannot exceed the smaller of the number of rows and the number of columns. Thus, the rank of X (and of X') is the smaller of d and n. Also, the rank of a matrix product is no greater than that of either factor (Perlis (1952), Ex. 7, p. 58). (The division by the scalar, d, to obtain A does not affect the rank.)

As a result, the rank of the covariance matrix A, which is always nxn, cannot exceed the smaller of n and d. Thus, if there are 40 locations, but only one day of observations, the rank of A would be 1. If there are 27 days of observations, the rank of A would not exceed 27. If there are 59 days of observations, the rank of A would not exceed 40. If there were 1241 days of observations, the rank of A would not exceed 40.

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA-650/4-75-005	2.	3. RECIPIENT'S ACCESSION NO.
4. TITLE AND SUBTITLE OBJECTIVE PROCEDURES FOR OPTIMUM LOCATION OF AIR POLLUTION OBSERVATION STATIONS		5. REPORT DATE June 1975
		6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) C. Eugene Buell		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS Kaman Sciences Corporation, P. O. Box 7463, Colorado Springs, Colorado 80933		10. PROGRAM ELEMENT NO. 1AA009
		11. CONTRACT/GRANT NO. EPA-68-02-0699
12. SPONSORING AGENCY NAME AND ADDRESS Environmental Sciences Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina 27711		13. TYPE OF REPORT AND PERIOD COVERED Final
		14. SPONSORING AGENCY CODE EPA-ORD
15. SUPPLEMENTARY NOTES		

16. ABSTRACT

This document is concerned with developing linear regression techniques for interpolation of air pollutant concentrations over an area and, using these techniques in a computer program for determining the optimum location of air pollution observing stations. The general interpolation problem is surveyed and the advantages of using linear regression formulas as interpolation formulas are discussed. The case of observations containing errors of observation or effects of limited range of influence is emphasized. Since the use of linear regression methods depends on knowledge of the two-point correlation function for pollutant concentration measures, the construction of correlation coefficients from synthetic data is taken up. Attention is given to the estimation of residual variances or the effects of limited range of influence, using Factor Analysis. In extending these methods to a continuous formulation in integral equation form, the lack of accuracy in the integral equation solution is more important than the statistical significance of the data unless the residual variances are removed. If this is done, then the tests for accuracy and statistical significance are reconciled. If the user carefully handles the residual variances in constructing program input, difficulties encountered in code development are avoidable.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Air pollution Linear regression Interpolation Factor analysis Optimization		13B 12A

18. DISTRIBUTION STATEMENT RELEASE TO PUBLIC	19. SECURITY CLASS *(This Report)* UNCLASSIFIED	21. NO. OF PAGES 215
	20. SECURITY CLASS *(This page)* UNCLASSIFIED	22. PRICE

EPA Form 2220-1 (9-73)

INSTRUCTIONS

1. **REPORT NUMBER**
Insert the EPA report number as it appears on the cover of the publication.

2. **LEAVE BLANK**

3. **RECIPIENTS ACCESSION NUMBER**
Reserved for use by each report recipient.

4. **TITLE AND SUBTITLE**
Title should indicate clearly and briefly the subject coverage of the report, and be displayed prominently. Set subtitle, if used, in smaller type or otherwise subordinate it to main title. When a report is prepared in more than one volume, repeat the primary title, add volume number and include subtitle for the specific title.

5. **REPORT DATE**
Each report shall carry a date indicating at least month and year. Indicate the basis on which it was selected *(e.g., date of issue, date of approval, date of preparation, etc.)*.

6. **PERFORMING ORGANIZATION CODE**
Leave blank.

7. **AUTHOR(S)**
Give name(s) in conventional order *(John R. Doe, J. Robert Doe, etc.)*. List author's affiliation if it differs from the performing organization.

8. **PERFORMING ORGANIZATION REPORT NUMBER**
Insert if performing organization wishes to assign this number.

9. **PERFORMING ORGANIZATION NAME AND ADDRESS**
Give name, street, city, state, and ZIP code. List no more than two levels of an organizational hirearchy.

10. **PROGRAM ELEMENT NUMBER**
Use the program element number under which the report was prepared. Subordinate numbers may be included in parentheses.

11. **CONTRACT/GRANT NUMBER**
Insert contract or grant number under which report was prepared.

12. **SPONSORING AGENCY NAME AND ADDRESS**
Include ZIP code.

13. **TYPE OF REPORT AND PERIOD COVERED**
Indicate interim final, etc., and if applicable, dates covered.

14. **SPONSORING AGENCY CODE**
Leave blank.

15. **SUPPLEMENTARY NOTES**
Enter information not included elsewhere but useful, such as: Prepared in cooperation with, Translation of, Presented at conference of, To be published in, Supersedes, Supplements, etc.

16. **ABSTRACT**
Include a brief *(200 words or less)* factual summary of the most significant information contained in the report. If the report contains a significant bibliography or literature survey, mention it here.

17. **KEY WORDS AND DOCUMENT ANALYSIS**
(a) DESCRIPTORS - Select from the Thesaurus of Engineering and Scientific Terms the proper authorized terms that identify the major concept of the research and are sufficiently specific and precise to be used as index entries for cataloging.

(b) IDENTIFIERS AND OPEN-ENDED TERMS - Use identifiers for project names, code names, equipment designators, etc. Use open-ended terms written in descriptor form for those subjects for which no descriptor exists.

(c) COSATI FIELD GROUP - Field and group assignments are to be taken from the 1965 COSATI Subject Category List. Since the majority of documents are multidisciplinary in nature, the Primary Field/Group assignment(s) will be specific discipline, area of human endeavor, or type of physical object. The application(s) will be cross-referenced with secondary Field/Group assignments that will follow the primary posting(s).

18. **DISTRIBUTION STATEMENT**
Denote releasability to the public or limitation for reasons other than security for example "Release Unlimited." Cite any availability to the public, with address and price.

19. & 20. **SECURITY CLASSIFICATION**
DO NOT submit classified reports to the National Technical Information service.

21. **NUMBER OF PAGES**
Insert the total number of pages, including this one and unnumbered pages, but exclude distribution list, if any.

22. **PRICE**
Insert the price set by the National Technical Information Service or the Government Printing Office, if known.