N 69 27946

NASA CR 61282

NASA CONTRACTOR
REPORT

Report No. 61282

MATHEMATICAL WIND PROFILES

PARTS I, II, III, and IV

By Arnold Court, Robert R. Read, and
Gerald E. Abrahms

Lockheed-California Company

May 20, 1969

NASA-George C. Marshall Space Flight Center
Marshall Space Flight Center, Alabama 35812

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No. NASA CR-61282	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle MATHEMATICAL WIND PROFILES, Parts I, II, III, IV		5. Report Date May 20, 1969
		6. Performing Organization Code
7. Author(s) Arnold Court, Robert R. Read, and Gerald E. Abrahms		8. Performing Organization Report No.
9. Performing Organization Name and Address Lockheed-California Company		10. Work Unit No.
		11. Contract or Grant No. NAS 8-5380
12. Sponsoring Agency Name and Address NASA-George C. Marshall Space Flight Center Marshall Space Flight Center, Ala. 35812		13. Type of Report and Period Covered Compilation of Parts I, II, III, and IV
		14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

SEE REVERSE SIDE OF THIS PAGE

17. Key Words	18. Distribution Statement Public Release		
19. Security Classif. (of this report) U	20. Security Classif. (of this page) U	21. No. of Pages 105	22. Price

ABSTRACT

Augmented complex Fourier polynomials, in which constant and linear terms are added to a Fourier series for a complex variate, are developed to represent a hodograph. They are applied to hodographs representing the projection, on a horizontal plane, of the wind at equal intervals in the vertical, and hence to describe the vertical profile of horizontal wind velocity. Reasons for selecting this function, and methods for its computation and application, are given in Part I, with polynominal coefficients for mean monthly winds over Cape Kennedy, Florida, and for four consecutive soundings over Montgomery, Alabama. In Part II, seasonal differences, differences between the velocity and momentum representations of the wind, length and specific interval in a sounding, and the effect of averaging and normalizing on a profile are discussed, with 11 tables based on two 34-km soundings and two sequences of 6-hourly soundings from Montgomery, Alabama. In Part III, serially complete 6-hourly wind observations from the surface to 27-km over Cape Kennedy during 1962 are used to compute 5200 serial correlations of wind integrated over 7-km layers. Criteria for wind profile predicting are formulated from these correlations, for four representations of the wind, four atmospheric zones and their sum, thirteen calendar intervals, and four time lags. In Part IV, properties of the augmented complex Fourier polynomials are summarized, and detailed procedures are developed for predicting a future vector wind profile from present and past profiles, for application to winds at Cape Kennedy, Florida.

NOTE: The work documented in this report was prepared under the sponsorship of the Aerospace Environment Division, Aero-Astrodynamics Laboratory, Marshall Space Flight Center, NASA, Huntsville, Alabama. Mr. Orvel E. Smith, Chief, Terrestrial Environment Branch, was the contract technical monitor.

The four parts of this Report previously received very limited distribution as Lockheed Reports 17683 (March 1964), 18734 (May 1965), 18989 (August 1965), and 19128 (October 1965). Part I was issued as NASA CR-61195 in February 1968, and is now rescinded by this combined version. The four parts have been edited slightly by the senior author to form four chapters of a single Report, with consecutive paging and unified references. Distribution is provided in the interest of information exchange; responsibility for the contents resides with the original authors.

TABLE OF CONTENTS

Section		Page
	LIST OF TABLES	v
I-1	INTRODUCTION (I)	1
I-2	REPRESENTATION	3
I-3	FUNCTIONS	5
I-4	SERIES	7
I-5	FOURIER	9
I-6	PROPERTIES	11
I-7	APPLICATIONS	14
I-8	DISCUSSION	17
	APPENDIX I-A	24
	ESTIMATION OF COEFFICIENTS	
	APPENDIX I-B	25
	ORTHOGONALITY	
II-1	INTRODUCTION (II)	29
	1.1 Model	29
	1.2 Harmonics	
	1.3 Applications	32
II-2	SCOPE OF WORK AND BASIC DATA	33
	2.1 Explained Variance	33
	2.2 Ranking Order	37
	2.3 Effect of Arbitrary Selection	39
	2.4 Long Profiles	41
II-3	SIMILARITY OF COMPLETE INTERVAL REPRESENTATIONS	43
II-4	SEQUENTIAL SIX-HOUR SOUNDINGS	48
II-5	MODIFICATION OF OBSERVATIONS	50
II-6	CONCLUSIONS (II)	51
	APPENDIX II-A	54
	LONG SOUNDINGS FOR MONTGOMERY, ALA.	

TABLE OF CONTENTS (Continued)

Section		Page
III-1	INTRODUCTION (III)	55
III-2	CORRELATION COMPUTATIONS	58
III-3	CORRELATION SIGNIFICANCE	62
III-4	RESULTS	66
III-5	COMBINED INTERVALS	67
III-6	CONCLUSIONS (III)	70
	APPENDIX III-A	71
	SERIAL CORRELATIONS FOR COMBINED INTERVALS	
IV-1	INTRODUCTION (IV)	81
IV-2	BASIC FORMULAS	82
IV-3	VARIANCE REDUCTION	84
IV-4	PREDICTION	85
IV-5	FORMULATION	87
IV-6	COMPUTATION	89
IV-7	CONCLUSIONS (IV)	91
	APPENDIX IV-A	93
	COMPLEX FOURIER ANALYSIS PROGRAM	
	APPENDIX IV-B	100
	PREDICTION EQUATIONS	
	APPENDIX IV-C	103
	PREDICTION AND VERIFICATION	
	REFERENCES	105

LIST OF TABLES

Table		Page
I-1	Coefficients of Augmented Fourier Polynomials, and Cumulative Reduction in Relative Variance, r^2, for Mean Monthly Winds Over Cape Kennedy, Florida, 1956-1961	19
I-2	Coefficients of Augmented Fourier Polynomials, and Cumulative Reduction in Relative Variance, r^2, for Wind Speed and Momentum Density over Montgomery, Alabama, 9-10 January 1956	20
II-1	Cumulative Percent Variance Explained (CPVE) by Linear, Linear Plus Four, and Linear Plus Six Highest Harmonics, Respectively, for Two Soundings at Montgomery, Alabama	35
II-2	Maximum, Minimum, and Range of CPVE from Table 1.	36
II-3	First Six Ranked Harmonic Numbers, on Basis of Percent Variance Explained	38
II-4	Frequency of Occurrence of the Four Most Important Harmonics Numbers	39
II-5	Difference in CPVE by Preselected and by Actually Best Harmonics, Given as Percent of Total Variance Explained for 4 and 6 Harmonic Terms	40
II-6	CPVE for Two Long Profiles, for Velocity and Momentum	41
II-7	Ranked Harmonic Numbers for Velocity and Momentum	42
II-8	W and r_{sav} Values	45
II-9	Rank Correlation Between (0) and (2) Intervals with All Other Intervals	47
II-10	CPVE for Consecutive Six-Hour Soundings	49
II-11	Frequency of Rank Correlations for Monthly Cape Kennedy Data	51
III-1	Average Monthly Density (g/cm^3) for December at Cape Kennedy	57
III-2	Calendar Intervals Used for Serial Correlations	58
III-3	First and Last Multiplicative Pair in the Cross-Product Term for the First Calendar Interval	61
III-4	95% Confidence Intervals for Correlation Coefficients Based on 28 and 112 Pairs	64

LIST OF TABLES (Continued)

Table		Page
III-5	Serial Correlations for 6 Hr. Lag, 0-27 Kilometers	76
III-6	Serial Correlations for 6 Hr. Lag, 0-6 Kilometers	77
III-7	Serial Correlations for 6 Hr. Lag, 7-13 Kilometers	78
III-8	Serial Correlations for 6 Hr. Lag, 14-20 Kilometers	79
III-9	Serial Correlations for 6 Hr. Lag, 21-27 Kilometers	80

MATHEMATICAL WIND PROFILES

PART I

SUMMARY

Augmented Fourier polynomials, in which constant and linear terms have been added to a complex Fourier series, appear to offer a means for representing the vertical profile of the horizontal wind velocity. Reasons for selecting this function, and methods of its computation and application, are given. Polynomial coefficients are presented for mean monthly winds over Cape Kennedy, Florida, and for four consecutive soundings over Montgomery, Alabama.

1. Introduction (I)

Mathematical representation of the vertical profile of wind is desirable for many purposes, and essential for the rigorous comparison of profiles and the prediction of profiles by statistical regression techniques. Because wind is a two-dimensional vector (neglecting the vertical component, which is at least an order of magnitude smaller than the horizontal components), the vertical profile of the instantaneous wind is a curve in three-dimensional space. The graphical and analytical difficulties in describing such a curve have thus far prevented any systematic description of complete wind profiles. In this report, various possible methods of representation are explored, and one of them, using complex Fourier series, is developed in detail. Application of the method, and its evaluation, will be the subjects of future reports.

Notation has been chosen carefully for consistency and clarity. The wind speed toward the east is denoted by x, that toward the north by y. Their vector resultant is called $\underset{\sim}{z}$, and the modulus or absolute value of the resultant is z:

$$|\underset{\sim}{z}|^2 = z^2 = x^2 + y^2. \tag{1.1}$$

The direction of this resultant, in degrees clockwise from north, is

$$\theta = \text{arc sin } \frac{x}{z} = \text{arc cos } \frac{y}{z} . \tag{1.2}$$

This double definition eliminates the ambiguity of sign inherent in a definition based on arc tan y/x. The meteorological convention for angles, used also in surveying and navigation, differs from the mathematical practice, in which angles are measured counterclockwise from the x-axis (east in meteorological practice). For the mathematical development, therefore, the direction is designated as

$$\phi = \frac{\pi}{2} - \theta = \text{arc sin } \frac{y}{z} = \text{arc cos } \frac{x}{z}, \tag{1.3}$$

and hence measured counterclockwise from east.

Alternative to the Cartesian (x, y), polar (z, θ), and vector ($\underline{z}$) representations of a wind vector is its representation as a complex variable, ζ:

$$\zeta = \underline{z} = x + i\,y = ze^{i\phi}. \tag{1.4}$$

To reduce the number of subscripts, a second wind vector will be denoted as (u, v), (w, ψ), $\underline{w}$, or $\eta = w \exp(i\xi)$. Height upward from the ground will be designated as h, atmospheric density as q, true correlation as ρ and its sample estimate as r, true variance as σ^2 and its sample estimate as s^2, and gravity as g.

The complex conjugate of a complex number will be denoted by an asterisk:

$$\zeta^* = x - iy = ze^{-i\phi}. \tag{1.5}$$

Therefore, the real and imaginary parts of the complex number ζ are

$$\begin{aligned} \tilde{\tilde{R}}(\zeta) &= \frac{\zeta + \zeta^*}{2} = z\,\frac{e^{i\phi} + e^{-i\phi}}{2} = z\cos\phi = x, \\ \tilde{\tilde{C}}(\zeta) &= \frac{\zeta - \zeta^*}{2} = z\,\frac{e^{i\phi} - e^{-i\phi}}{2} = z\sin\phi = y. \end{aligned} \tag{1.6}$$

Other notation will be identified when used.

2. Representations

Because a wind profile is a curve in three-dimensional space, its graphical representation on two-dimensional paper requires elimination of one dimension. Various graphical methods have been used for many years, each with some advantages and many disadvantages. The four basic methods, illustrated in Figure 1 with mean January winds for Cape Kennedy, Florida, are

a. each component, separately, vs height

b. speed and direction, separately, vs height

c. velocity hodograph

d. position hodograph.

The first two methods require mental addition of values from the two lines to give a picture of the actual wind vector and its changes. This difficulty is eliminated in the hodographs, in which the vertical dimension (or time) is indicated only by successive points along the path.

A hodograph is a curve connecting the end-points of successive vectors drawn from a common origin. The vectors may be successive in height, to represent the wind profile, or in time, to show the time variation of wind. The former application is used here, but the mathematical formulation is equally applicable to the time series case. The vectors may represent the actual wind velocity at each level, or they may represent the integral of the velocity, which gives the position of an object, such as a balloon, rising with constant speed through the wind field. The usual plotting-board representation of a pilot balloon trajectory is a position hodograph of the vertical wind profile, while the similarity trajectory of a constant-level balloon is a position hodograph of the time variation of wind. A position hodograph can be prepared from wind velocity information by plotting the successive vectors additively rather than from a common origin.

Hodographs appear more suitable for mathematical representation of the vertical wind profile than separate representations by components, or by speed and direction. But choice between the two hodographs, velocity and position, is more difficult. Fortunately, the computational procedures of fitting a function to observations are the same for either type of hodograph, since the purpose is merely to obtain an analytic function describing the curve.

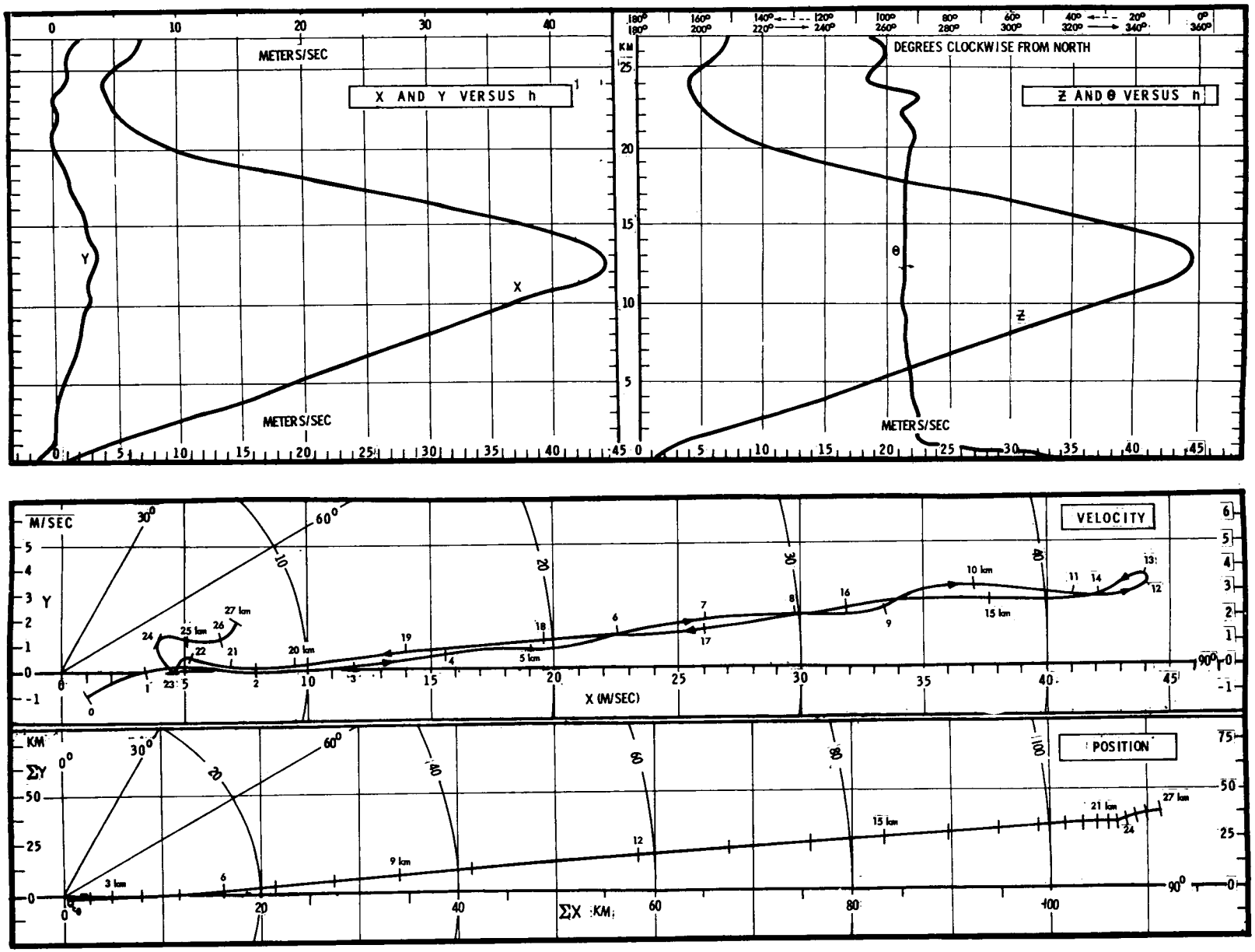
METERS/SEC
X AND Y VERSUS h
Y
X
METERS/SEC
KM
DEGREES CLOCKWISE FROM NORTH
Ż AND Θ VERSUS h
Θ
Ż
METERS/SEC
M/SEC
VELOCITY
Y
X (M/SEC)
KM
POSITION
ΣY
ΣX :KM

FIGURE 1

When positions actually are measured (as in most meteorological observations using balloons, rising or falling), the position hodograph should be fitted. One differentiation of the fitted function then will give the velocity hodograph function, and a second differentiation the wind shear, which is of considerable importance. Actually, most routine wind information is obtained from finite differences of balloon positions, and shears from finite differences of these computed velocities, i.e., by smoothed second differences of the basic observations.

When wind velocities are obtained directly, as by sound ranging, the velocity hodograph should be fitted. One differentiation then will yield shears, while integration gives the positions to which they apply. Such positional information is needed for studies of the trajectories of falling or suspended objects, such as radioactive fallout or toxic pollutants.

Any mathematical function used to approximate a hodograph must be continuous and have continuous first and second derivatives. Since the hodograph is a vector-valued function $\underline{z}(h)$ of a scalar argument, h, in practice, representation by components is more convenient. Compactness of representation and relative ease of manipulation make the complex form,

$$x(h) + iy(h), \qquad z(h) \exp [i\phi(h)],$$

suited for an attempt at developing an expression for $\underline{z}(h)$.

3. Functions

Selection of a mathematical function to approximate the vertical wind profile, as represented by its position or velocity hodograph, must be based largely on convenience and general suitability, including possession of continuous derivatives. Meteorological theory and hydrodynamic theory are as yet inadequate to provide a definitive functional form, except for certain height ranges.

In the lowermost ten meters of the atmosphere, air flow increases with height without material change in direction (Hess, 1959) [1]. When the temperature lapse rate is neutral, the logarithmic wind profile appears to fit available observations:

$$z = \frac{\sqrt{\tau/\rho}}{k} \ln \frac{h}{h_o} \qquad (3.1)$$

where τ is the eddy stress, q the density, k von Karman's constant, and h_o a "roughness parameter." When the lapse rate is not neutral, an exponential profile seems more appropriate:

$$z = z_1(h/h_1)^m \tag{3.2}$$

where z_1 is the wind speed at height h_1 (usually a few centimeters) and m is a positive exponent less than unity. A generalization, for variable lapse rates, is offered by the Deacon profile:

$$z = \left[\frac{\sqrt{\tau/q}}{k(1 - \beta)}\right]\left[\left(\frac{h}{h_o}\right)^{1-\beta} - 1\right]. \tag{3.3}$$

For several hundred meters above this boundary layer, wind increases in speed with height, and turns clockwise, in the northern hemisphere, generally according to the Ekman spiral. At about the 10-meter level, the wind is directed toward the left of the geostrophic wind, which blows along the isobars at 1 km or higher. The wind vector at height h in this spiral or friction layer is

$$z(h) = z_g\left[e^{i\phi} - e^{-ah}\, e^{i(ah-\phi)}\right]. \tag{3.4}$$

Here z_g is the magnitude of the geostrophic wind, blowing at an angle ϕ (in mathematical notation) to the positive x-axis, and a is a function of density, Coriolis force, and eddy viscosity. Actual winds do follow this Ekman spiral when the upper wind flow is straight or only slightly curved, and the lowermost kilometer of air has no appreciable horizontal gradients of temperature.

Above the spiral layer, wind speed generally increases with height up to the level of maximum wind, which usually occurs slightly below the tropopause at 10 to 12 km. Often the increase in speed with height is at about the same rate as the decrease of density with height, so that between 5 and 10 km "Egnell's law" states that the momentum is constant. A justification of this empirical rule, deduced from cloud and pilot balloon observations 70 years ago by Clayton in Massachusetts and Egnell in France, was offered by Humphreys (1929, pp. 135-136) [**2**].

Above the maximum wind layer, wind speed decreases with height to a minimum, on the average, at 22 to 25 km, but no law or rule describing this decrease, or the accompanying change in direction, has yet appeared. Thus, while some theoretical formulations are available for wind behavior in the boundary and spiral layers, a few guidelines can be found for the form of a function to describe the wind profile above 1 km.

4. Series

In the absence of any theory on which to base a functional form for wind profile description, some empirical function must be chosen. Logical candidates for this purpose are polynomials. The wind vector $\underline{z} = (x, y)$ could be represented as a function of height, h, by two separate polynomials, one for each component:

$$x_{h,m} = \sum_{k=0}^{m} a_k h^k, \qquad y_{h,n} = \sum_{k=0}^{n} b_k h^k \tag{4.1}$$

where m and n are the numbers of terms required for satisfactory fit or agreement of the polynomial with the observations. Agreement would be determined by the variance (mean squared difference) of the observations about the polynomials. The absolute or unconditional variances are, respectively, s_x^2 and s_y^2, and the conditional variances $s_{x,m}^2$ and $s_{y,n}^2$:

$$s_x^2 = \nu^{-1} \sum (x_h - \bar{x})^2 = \nu^{-1} \sum x_h^2 - (\bar{x})^2,$$

$$s_{x,m}^2 = \nu^{-1} \sum (x_h - x_{h,m})^2 = \nu^{-1} \sum x_h^2 + \nu^{-1} \sum x_{h,m}(x_{h,m} - 2x_h), \tag{4.2}$$

and similarly for s_y^2 and $s_{y,n}^2$. (All summations are for h = 0, 1, 2, ..., N, and $\nu = N + 1$.) The extent to which the variance of x is reduced by use of an m-term polynomial is

$$s_x^2 - s_{x,m}^2 = \nu^{-1} \sum x_{h,m} (x_{h,m} - 2x_h) - (\bar{x})^2. \tag{4.3}$$

Of greater interest than this absolute reduction in variance is the relative reduction, or squared correlation (sometimes called the coefficient of determination):

$$r^2_{x,m} = \frac{s^2_x - s^2_{x,m}}{s^2_x} = \frac{\sum x_{h,m}\,(x_{h,m} - 2x_h) - \nu(\bar{x})^2}{\sum x_h^2 - \nu(\bar{x})^2} . \qquad (4.4)$$

Similar expressions yield the absolute and relative reductions in the variance of y.

As more and more polynomial terms are used, i.e., as m and n increase, the variance reduction increases and the correlations approach one, attaining this value for $m = \nu = n$. But when $r^2 = .9$, the fit of the polynomial to the observations is considered adequate for most purposes, although in some cases values as high as .95 are desired.

However, the various terms of the polynomials may not be equally effective in reducing the variance. A higher power, such as a_4h^4, may be more effective than a lower one. Hence, the terms should be chosen not in simple order, but according to the amount of variance reduction that they provide.

A more efficient polynomial, in the sense of having fewer terms, would be formed from those terms, regardless of their exponents, providing the greatest reduction in variance, or highest correlation. The various terms, a_kh^k, should be arranged according to their contribution to the variance reduction. Coefficients ordered in this way may be denoted as $a_{(k)}h^{(k)}$, and the first m of them will be considered to form the index set M.

In this notation, the polynomial providing the required (e.g., 90%) relative reduction in variance is

$$x_{h,M} = \sum_{(k)=1}^{m} a_{(k)}h^{(k)} = \sum_{k\in M} a_k h^k , \qquad (4.5)$$

and similarly for $y_{h,N}$.

Such polynomials would provide suitably efficient procedures for representing each of the components separately. But they offer no link between the components; they do not apply to the wind vector itself. When results obtained by two such polynomials are combined to provide estimates of the wind vector at each level, excessive interlevel shears could be indicated. Hence, they do not seem particularly suited for the mathematical representation of wind vectors.

The same objections apply to the fitting of a complex variable by a single power series with complex coefficients:

$$\zeta_{h,M} = \sum_{k\epsilon M} c_k h^k = \sum_{k\epsilon M} (a_k + i\, b_k)\, h^k = \sum_{k\epsilon M} a_k h^k + i \sum_{k\epsilon M} b_k h^k. \qquad (4.4)$$

These objections to expressing the wind components as polynomial functions of height apply regardless of the method of estimating the polynomial coefficients. Orthogonal polynomials, while possessing the great advantage that they need not be recomputed after selection of the highest-order term contributing significantly to the variance reduction, are no better in these respects than simple power series.

5. Fourier

Complex trigonometric polynomials (Fourier series) are not subject to the same drawbacks as univariate polynomials, just discussed. The estimation of the coefficients of each component (i.e., the real and imaginary parts) is based on both components of the observed wind, and hence such a complex series actually estimates the vector, or entire complex number, rather than separate components.

Fourier series often are used to represent functions known to be periodic, but are not restricted to such use. Lighthill (1960) [3] declares (p.4) that a common application is "to represent a function which is not periodic, but instead is defined in the first place only in a restricted interval," covering perhaps 30 km in the vertical. Wind information usually is available only for a restricted interval. Description of the time and space variations in such a 30-km profile may be possible through the fitting of Fourier series or polynomials.

Such polynomials, however, have no linear terms. Since the wind often increases rather regularly with height, at least over certain height ranges, a linear term obviously is desirable in any expression for the vertical wind profile. This can be provided by defining a plane about which the actual wind observations vary, and then describing such

variations by Fourier polynomials. The required plane is defined by two intersecting straight lines, in the vertical x, h and y, h planes, respectively, that represent the individual wind components.

The original observations of the wind at level h,

$$\zeta_h = x_h + i\, y_h = z_h \exp(i\phi), \tag{5.1}$$

may be expressed in terms of the least squares linear trends as

$$\zeta_h = c_z + d_{oo}h + \eta_h. \tag{5.2}$$

The departure

$$\eta_h = u_h + i\, v_h$$

is given by

$$u_h = x_h - c_x - a_{oo}\, h, \quad v_h = y_h - c_y - b_{oo}h. \tag{5.3}$$

The linear coefficients - reasons for the double zero subscripts will be apparent later - are

$$a_{oo} = \frac{\nu \sum x_h\, h - \sum x_h \sum h}{\nu \sum (h - \bar{h})^2}, \quad b_{oo} = \frac{\nu \sum y_h\, h - \sum y_h \sum h}{\nu \sum (h - \bar{h})^2}. \tag{5.4}$$

The constant terms are

$$c_x = \bar{x} - a_{oo}\bar{h}, \quad c_y = \bar{y} - b_{oo}\bar{h}. \tag{5.5}$$

Thus, the variations of the wind vector about the least squares plane are

$$\eta_h = \zeta_h - (\bar{\zeta} - d_{oo}\bar{h}) - d_{oo}h, \tag{5.6}$$

where $d_{oo} = a_{oo} + i\, b_{oo}$ is obtained from (5.4).

Fourier polynomials describing η_h are

$$\eta_{h,M} = \sum_{j\in M} d_j \exp(i\lambda\, jh), \quad \lambda = 2\pi/\nu. \tag{5.7}$$

The complex coefficients $d_j = a_j + i\, b_j$ are estimated (as explained in Appendix A, and discussed in the next section) from the ν values of η_h, obtained from the ν observations of ζ_h. Summation is over the set M of the m terms contributing most to the reduction in variance, as discussed in the previous section for univariate polynomials.

After the $\{d_j\}$ have been estimated and the set M chosen, the resulting Fourier polynomial can be augmented by the constant and linear terms to provide a complete expression for the actual wind profile:

$$\zeta_{h,M} = \bar{\zeta} + d_{oo}(h - \bar{h}) + \sum_{j\in M} d_j \exp(i\lambda\, jh). \tag{5.8}$$

Application of this expression for the wind profile to actual wind observations is discussed in the following sections.

6. Properties

Expansion of (5.7) shows that the estimation of each component of the wind vector $\eta_{h,M}$ and hence of $\zeta_{h,M}$, involves coefficients from both the real and imaginary parts of the polynomial:

$$\eta_{h,M} = \sum_{j\in M} (a_j + i\, b_j)(\cos\lambda\, jh + i \sin\lambda\, jh) = \sum_{j\in M} (a_j \cos\lambda\, jh - b_j \sin\lambda\, jh)$$

$$+ i \sum_{j\in M} (b_j \cos\lambda\, jh + a_j \sin\lambda\, jh). \tag{6.1}$$

The least squares estimators of the complex coefficients d_j are, as shown in Appendix A,

$$d_j = a_j + i\, b_j = \frac{1}{\nu} \sum_{h=0}^{N} \eta_h \exp\,(-\, i\lambda\, jh)$$

$$= \frac{1}{\nu} \sum_{h=0}^{N} (u_h + i\, y_h)\, (\cos\, \lambda\, jh - i \sin\, \lambda\, jh) \qquad (6.2)$$

$$= \frac{1}{\nu} \sum_{h=0}^{N} (u_h \cos\, \lambda\, jh + v_h \sin\, \lambda\, jh) + i\, \frac{1}{\nu} \sum_{h=0}^{N} (v_h \cos \lambda\, jh - u_h \sin\, \lambda\, jh).$$

That these estimators actually minimize the sum of the squared departures of the observations from the least-squares regression plane is shown in Appendix A. These squared departures are the sums of the squared departures of the two components; divided by ν, the total number of observations, they yield the conditional variance about the polynomial:

$$\frac{1}{\nu}\, \sigma^2_{\eta;M} = S^2_{\eta;M} = \sum_{h=0}^{N} S^2_{\eta,h;M} = \sum_{h=0}^{N} (\eta_h - \eta_{h;M})\, (\eta_h - \eta_{h;M})^* \,. \qquad (6.3)$$

A major purpose of this study is to determine the magnitude of the absolute reduction in variance, $\sigma^2_\eta - \sigma^2_{\eta;M}$ and the relative reduction, $r^2_{\eta;M}$ (4.4), when a wind profile, from which ν observations are obtained at equal height intervals, is approximated by (5.8) for $m \leq 4$. If the representation is adequate, $\zeta_{h;M}$ may be evaluated for any value of h, not necessarily those equally-spaced values at which ζ_h was observed. This would provide a continuous representation of a wind profile originally described for discrete points only.

In addition, the function (5.8) can be differentiated to provide a continuous representation of the wind shear, $\partial\zeta_{h;M}/\partial_h$. Alternatively, the ζ_h may be the balloon positions at successive heights, and differentiation then will provide wind speeds at any height.

Not only do the coefficients $\{d_j\}$, estimated by (6.2), minimize $S^2_{\eta;M}$, but, as discussed in Appendix B, they seem to be approximately orthogonal, although the precise extent of any slight dependence between them is still to be determined.

Orthogonality insures that for any set $\{M\}$ of coefficients,

$$S^2_{\eta,M} = \sum S^2_{\eta,j},$$

that is, that the contribution of each term to the total variance does not depend on what other terms are included in that total. This desirable property has been assumed in the preliminary applications of Fourier polynomials to the description of wind profiles.

Orthogonality properties are increased when the original observations ζ_h, expressed as departures η_h from the least-squares plane, all have the same variance. Thus, rather than η_h as defined by (5.6), computations of d_j by (6.2) should use $\eta_h/\sigma_{\eta;h}$, where $\sigma^2_{\eta;h}$ is the variance η_h. Since η_h is, by (5.6), a linear function of ζ_h, their variances are the same. Such variances should be used, when available, to adjust the values of η_h, as just indicated.

When the original observations $\zeta_h = x_h + i\, y_h$ are means, as for a month or season, variances are available for such adjustment. But when they are single observations, the proper choice of values is not obvious. In the following sections, examples are given of profiles computed from mean values adjusted for variance, and of profiles fitted to individual sets of observations without variance adjustment. The propriety of this second procedure, although it seems to provide an adequate fit, requires further investigation.

Another topic for further study is the method of computing the plane about which the departures η_h are taken. The Fourier polynomials may provide an even better approximation to the observations if this trend plane is constructed through the mean point so that the first and last observations (lowest and highest wind observations) are equidistant from it.

7. Applications

Augmented Fourier polynomials, as developed in the preceding two sections, were fitted to two sets of wind data to determine whether the method showed sufficient promise to warrant further study and development. Results of such application, presented in this section, are quite encouraging.

One set of wind data was composed of monthly mean winds, at 1-km levels, over Cape Kennedy, Florida. They are based on 5 years of observations (the first 321 days were at nearby Patrick Air Force Base), 1956-1961. Missing observations had been interpolated before averaging, so that sample sizes were the same at all levels. These data were furnished by Mr. Orvel E. Smith of the Aero-Astrodynamics Laboratory, George C. Marshall Space Flight Center, in advance of publication.

The other set was made up of four consecutive observations, at 6-hour intervals, over Montgomery, Alabama, on 9 January 1956. These were the first four consecutive soundings, each reaching to at least 25 km, in an extensive compilation of winter and summer soundings furnished by the National Weather Records Center, U. S. Weather Bureau, at Mr. Smith's request. These soundings also contained data on atmospheric density, so that momentum density as well as wind speed could be fitted by augmented Fourier polynomials. (Units of momentum density, the product of wind speed and atmospheric density, are dynes per cubic centimeter.)

These two sets of data provided a total of 20 "soundings," each sounding being a set of values of ζ_h for successive values of h. Of these, 12 were monthly means for Cape Kennedy, four were successive wind observations at Montgomery, and four were the corresponding momentum density observations. For each such "sounding," the lowermost 2 km were ignored, because of possible friction layer effects, as discussed in Section 3, and only the levels from 2 to 25 km, inclusive, were used. In the notation already developed, $h_0 = 2$ km, $h_1 = 3$ km, ..., $h_N = 25$ km.

Results of the fitting of the augmented Fourier polynomials to these 20 soundings are given in Tables 1 and 2. After the constant and linear terms, the coefficients are presented in decreasing order of the amount of variance "explained" by them. That is, the coefficients d_j have been ordered as $d_{(j)}$, as discussed in Section 4. For example, in the first line of Table 1 (for January mean winds over Cape Kennedy), $a_{(1)}$ and $b_{(1)}$ are, respectively, a_{23} and b_{23}, so that j = 23 is used in the trigonometric terms that they multiply.

Coefficients are given in Tables 1 and 2 for each wind component separately, as indicated in the formulas at the head of Table 2, which are based on (5.8) and (6.1). The two formulas may be combined into one expression, in complex notation. Thus, the mean January wind over Cape Kennedy may be written as

$$\zeta^{s}_{h,M} = (2.61 + 0.126\ i) - (0.054 - 0.003\ i)\ h$$

$$- (0.575 - 0.014\ i) \cos 23\pi\ h/12 - (0.014 + 0.575\ i) \sin 23\pi\ h/12$$

$$- (0.530 + 0.100\ i) \cos \pi\ h/12 + (0.100 - 0.530\ i) \sin \pi\ h/12$$

$$+ (0.044 + 0.173\ i) \cos 22\pi\ h/12 - (0.173 - 0.44\ i) \sin 22\pi\ h/12$$

$$+ (0.043 - 0.140\ i) \cos 2\pi\ h/12 + (0.140 + 0.043\ i) \sin 2\pi\ h/12. \qquad (7.1)$$

The superscript "s" indicates that the values of $\zeta^{s}_{h,M}$ obtained from (7.1), and from Table 1 generally, are for "standardized" values. They must be multiplied by the standard deviations of the wind components for the appropriate level to give values approximating the observed means.

For example, evaluation of (7.1) for h = 10, i.e., 12 km, gives 2.41 + 0.226 i. When each of these values is multiplied by the standard deviation of the corresponding wind component at 12 km over Cape Kennedy in January, 16.04 and 14.24 m/sec, respectively, estimated wind speeds are obtained which may be compared with the observed means:

Estimated	$x_{10} = 38.66$	$y_{10} = 3.22$
Observed	44.04	3.26.

In Figure 2, five hodographs are shown for the mean January winds over Cape Kennedy. In the upper panel, one hodograph depicts the actual means, in meters per second, while a second one shows the effect of dividing the speed at each level by its standard deviation, and expressing the result as a departure from the least-squares plane. The "trend" hodograph is centered at the origin, and is in units much smaller than those of the original values.

The lower panel of Figure 2 shows three hodographs, computed by Fourier polynomials, not augmented, i.e., as variations about the least-squares plane. The "one-term" hodograph is a circle, representing only the j = 23 term, without the preceding constant and linear terms or the

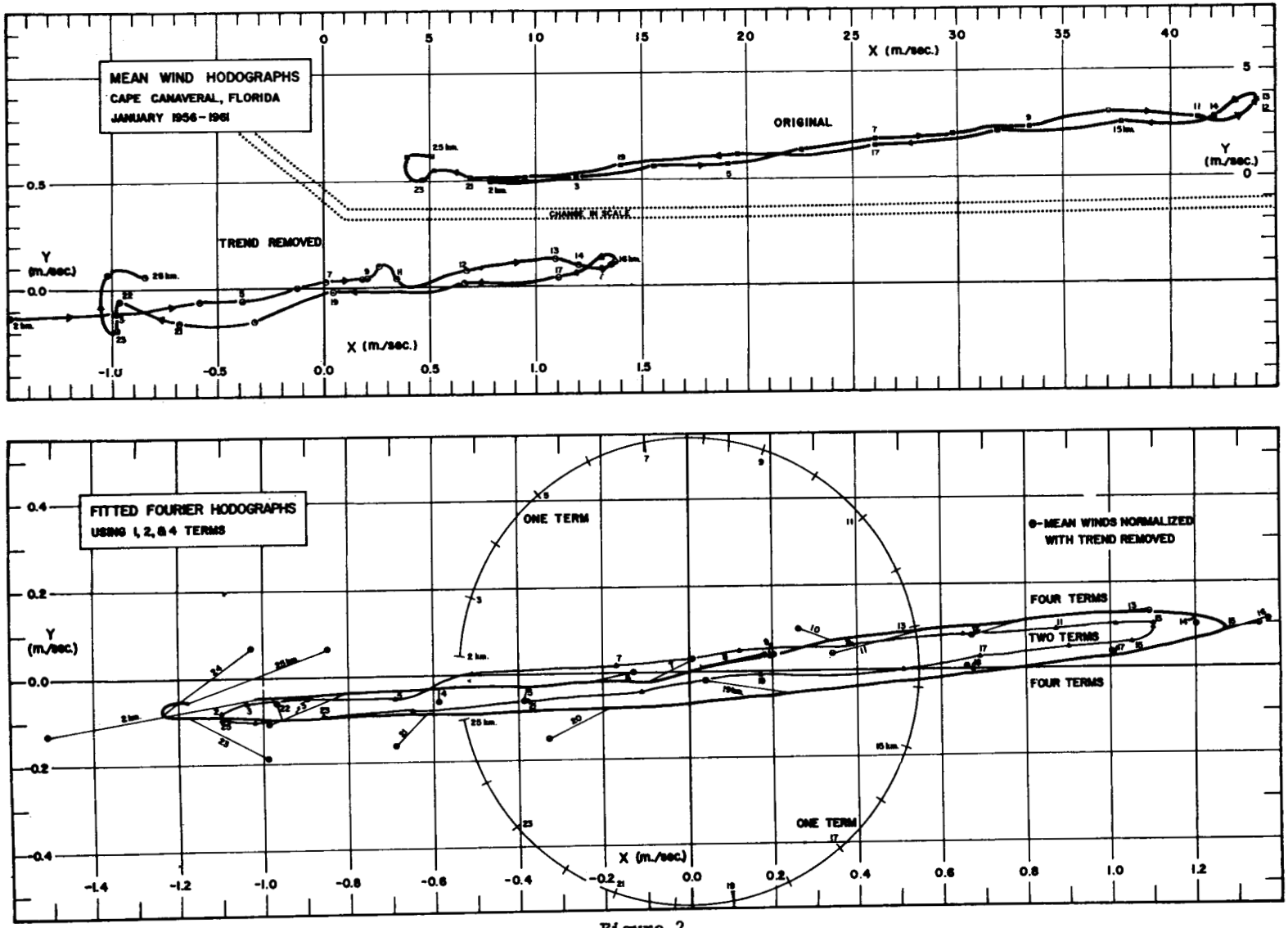
MEAN WIND HODOGRAPHS
CAPE CANAVERAL, FLORIDA
JANUARY 1956 - 1961
ORIGINAL
TREND REMOVED
CHANGE IN SCALE
X (m./sec.)
Y (m./sec.)
FITTED FOURIER HODOGRAPHS
USING 1, 2, & 4 TERMS
ONE TERM
FOUR TERMS
TWO TERMS
FOUR TERMS
ONE TERM
e-MEAN WINDS NORMALIZED WITH TREND REMOVED

Figure 2

final three terms. The "two-term" hodograph represents computation of the j = 23 and j = 1 terms in (7.1), without the constant and linear terms or the final two terms. The "four-term" hodograph presents results of using all terms of (7.1) except the constant and linear.

Shown as dots in the lower panel of Figure 2 are the same points, for each 1-km level, as in the "trend removed" hodograph of actual winds in the upper panel. The thin lines from these dots to the "four-term" curve indicate the extent of the vector difference between the observed mean winds, at each level, and the values computed from (7.1). The sum of the squares of the lengths of these thin lines is the $S^2_{\eta,M}$ of (6.3), for M = 4.

For the individual soundings over Montgomery, no estimates of wind variance at each level were readily available. The observed values were assumed to have the same variance, and no adjustments were made. Thus, the coefficients in Table 2, when introduced into the appropriate formula, give estimated winds directly in meters per second.

8. Discussion

Under each pair of coefficients in Tables 1 and 2 are two additional entries: the value of the index j for the pair, and the value of r^2, the relative reduction in variance (4.4) attained by using that term, and all preceding ones, in the augmented Fourier polynomial.

For the Cape Kennedy mean monthly wind profiles, the constant and linear terms alone reduce the variance by 80 percent in summer, but hardly at all in November and December. Two additional terms provide r^2 of 85 percent or more in all months, indicating that augmented Fourier polynomials of as few as four terms (m = 2) may provide descriptions adequate for some purposes. In nine of the months, term 23 provides the greatest reduction in variance, followed by term 1, while the same terms appear in reverse order in the other three months.

For all four Montgomery 6-hourly soundings, term 1 contributes most to the reduction in variance for both wind speed and momentum density. But whereas term 23 is second most important for wind speed, terms 2 (once) and 22 (thrice) have this role for momentum density. Values of r^2 for momentum density are consistently higher than for wind speed alone. Most of this difference arises in the constant and linear terms, for which r^2 is between 75 and 86 percent for momentum density, but only from 39 to 44 percent for wind speed. This may be a reflection of "Egnell's law," outlined in Section 3, and requires further study.

The extent to which these results depend on the particular height interval chosen also requires additional investigation. The strongest wind speeds in all the soundings are near the middle of the 2 to 26 km interval studied, which may explain the consistent appearance of term 1 as contributing significantly to the relative reduction in variance. Similarly, the importance of term 23 may indicate excessive level-to-level variability, perhaps actual but also possibly arising from observational errors and computational procedures in the compilation of wind information.

These and other considerations indicate that the most fruitful application of augmented Fourier polynomials to wind profile description may be their use to describe the position hodograph, as obtained directly from a balloon or other indicator, and the subsequent differentiation of the polynomial to provide wind speeds. This may provide considerable improvement over the present method employing successive finite differences, and may give greater detail of the wind profile and of its derivative, the wind shear.

Other topics for further study are statistical tests for the similarity or differences of two wind profiles, leading to criteria for their combination. For example, are January and February wind profiles over Cape Kennedy sufficiently similar that a combined winter profile describes them adequately? Also requiring study are procedures for predicting one profile from another, as in the case of the 6-hourly soundings over Montgomery.

Despite the need for these various extensions of the study, and further elaboration of the technique, the work reported here shows that mathematical description of an entire wind profile, either means or "instantaneous," can be attained with acceptable precision by the use of augmented Fourier polynomials.

TABLE 1. COEFFICIENTS OF AUGMENTED FOURIER POLYNOMIALS, AND CUMULATIVE REDUCTION IN RELATIVE VARIANCE, r^2, FOR MEAN MONTHLY WINDS OVER CAPE KENNEDY, FLORIDA, 1956-1961

	c_x	c_y	a_{oo}	b_{oo}	$a_{(1)}$	$b_{(1)}$	$a_{(2)}$	$b_{(2)}$	$a_{(3)}$	$b_{(3)}$	$a_{(4)}$	$b_{(4)}$
JAN	2.61	0.126	-0.054	0.003	-0.575	0.014	-0.530	-0.100	0.044	0.173	0.043	-0.140
	j=00		r^2=16.5		j=23	r^2=55.6	j=1	r^2=90.0	j=22	r^2=93.7	j=2	r^2=96.2
FEB	2.11	0.409	-0.048	-0.017	-0.505	-0.039	-0.492	-0.073	0.041	0.181	0.077	-0.168
	j=00		r^2=17.2		j=23	r^2=53.1	j=1	r^2=87.8	j=22	r^2=92.6	j=2	r^2=97.4
MAR	2.90	0.320	-0.086	-0.014	-0.695	-0.173	-0.675	0.052	0.074	-0.103	-0.057	0.085
	j=00		r^2=26.1		j=1	r^2=62.8	j=23	r^2=95.7	j=2	r^2=96.9	j=3	r^2=97.6
APR	1.69	-0.055	-0.056	-0.010	-0.548	0.007	-0.509	0.091	0.050	-0.138	0.046	0.128
	j=00		r^2=19.9		j=1	r^2=58.5	j=23	r^2=92.9	j=2	r^2=95.7	j=22	r^2=98.1
MAY	1.43	0.190	-0.098	-0.022	-0.642	0.050	-0.540	-0.034	0.101	0.155	-0.037	-0.132
	j=00		r^2=38.4		j=23	r^2=71.5	j=1	r^2=94.8	j=22	r^2=97.5	j=2	r^2=99.0
JUN	1.67	-0.035	-0.202	-0.021	-0.610	0.234	-0.584	0.104	0.149	-0.080	-0.021	-0.121
	j=00		r^2=70.1		j=23	r^2=85.2	j=1	r^2=97.7	j=1	r^2=98.7	j=2	r^2=99.2
JUL	1.41	0.298	-0.277	-0.036	-0.513	0.282	-0.545	0.097	0.120	-0.001	0.073	-0.062
	j=00		r^2=84.0		j=23	r^2=91.7	j=1	r^2=98.6	j=22	r^2=98.9	j=2	r^2=99.1
AUG	1.46	0.208	-0.263	-0.023	-0.579	0.237	-0.589	-0.026	0.170	0.029	0.149	-0.042
	j=00		r^2=79.8		j=23	r^2=89.2	j=1	r^2=97.5	j=22	r^2=98.2	j=2	r^2=98.8
SEP	1.06	0.174	-0.157	-0.018	-0.584	0.139	-0.558	-0.108	0.093	-0.012	0.058	-0.063
	j=00		r^2=62.0		j=23	r^2=80.7	j=1	r^2=97.4	j=22	r^2=97.9	j=2	r^2=98.3
OCT	1.43	0.064	-0.075	-0.013	-0.435	0.153	-0.420	-0.079	-0.043	-0.083	-0.033	0.079
	j=00		r^2=38.9		j=23	r^2=69.0	j=1	r^2=94.8	j=21	r^2=96.0	j=3	r^2=97.0
NOV	1.29	-0.064	-0.011	0.000	-0.332	0.113	-0.325	-0.067	-0.043	0.078	-0.022	-0.060
	j=00		r^2=2.2		j=23	r^2=47.9	j=1	r^2=88.8	j=3	r^2=91.7	j=21	r^2=93.2
DEC	1.63	0.234	0.009	0.002	-0.366	-0.104	-0.372	0.027	0.054	-0.090	0.008	0.098
	j=00		r^2=1.2		j=1	r^2=43.8	j=23	r^2=84.7	j=2	r^2=87.9	j=22	r^2=90.8

TABLE 2. COEFFICIENTS OF AUGMENTED FOURIER POLYNOMIALS, AND CUMULATIVE REDUCTION IN RELATIVE VARIANCE, r^2, FOR WIND SPEED AND MOMENTUM DENSITY OVER MONTGOMERY, ALABAMA, 9-10 JANUARY 1956

Wind components, x_h and y_h, at height h (in km starting at 2 km MSL), where $H = \pi h/12$, given by:

$$x_h = (c_x + a_{oo}h) + (a_{(1)} \cos j_1H - b_{(1)} \sin j_1H) + (a_{(2)} \cos j_2H - b_{(2)} \sin j_2H) + (a_{(3)} \cos j_3H - b_{(3)} \sin j_3H) + \ldots$$

$$y_h = (c_y + b_{oo}h) + (a_{(1)} \sin j_1H + b_{(1)} \cos j_1H) + (a_{(2)} \sin j_2H + b_{(2)} \cos j_2H) + (a_{(3)} \sin j_3H + b_{(3)} \cos j_3H) + \ldots$$

HOUR	c_x c_y	a_{oo} b_{oo}	$a_{(1)}$ $b_{(1)}$	$a_{(2)}$ $b_{(2)}$	$a_{(3)}$ $b_{(3)}$	$a_{(4)}$ $b_{(4)}$
			Wind Speeds			
0900	29.04 -33.62 j=00	-1.046 1.365 r^2=44.1	-8.852 -3.035 j=1 r^2=71.4	-5.542 3.716 j=23 r^2=85.3	-0.016 -3.253 j=22 r^2=88.6	0.079 2.779 j=2 r^2=91.0
1500	22.95 -27.52 j=00	-0.903 1.050 r^2=38.6	-7.843 -0.571 j=1 r^2=67.6	-4.743 4.604 j=23 r^2=82.9	-2.465 -0.154 j=22 r^2=85.5	1.260 1.808 j=2 r^2=87.5
2100	26.46 -26.42 j=00	-1.016 1.052 r^2=41.3	-8.115 -1.507 j=1 r^2=68.8	-6.021 3.499 j=23 r^2=88.4	1.840 2.610 j=2 r^2=92.5	-1.681 0.566 j=17 r^2=93.8
0300	28.42 -22.88 j=00	-1.146 0.913 r^2=42.9	-7.428 -1.161 j=1 r^2=66.5	-3.612 4.576 j=23 r^2=80.7	2.872 1.630 j=2 r^2=85.3	-0.697 -2.806 j=22 r^2=88.8
			Momentum Density			
0900	15.14 -22.92 j=00	-0.705 1.231 r^2=80.3	-1.036 -3.499 j=1 r^2=91.4	-0.943 1.357 j=2 r^2=93.7	-0.504 -1.321 j=22 r^2=95.4	-0.957 -0.687 j=23 r^2=96.6
1500	11.16 -17.63 j=00	-0.509 0.916 r^2=74.4	-1.293 -2.339 j=1 r^2=84.5	-1.572 -0.409 j=22 r^2=88.2	-0.832 -0.750 j=19 r^2=90.0	-0.663 0.630 j=3 r^2=91.2
2100	13.57 -17.71 j=00	-0.638 0.942 r^2=83.4	-0.826 -2.474 j=1 r^2=92.5	-0.505 -0.776 j=22 r^2=93.7	0.092 0.914 j=2 r^2=94.8	-0.800 -0.363 j=23 r^2=95.8
0300	15.70 -14.69 j=00	-0.773 0.768 r^2=86.3	-0.543 -1.739 j=1 r^2=91.3	-0.939 -0.720 j=22 r^2=93.4	-0.894 -0.717 j=21 r^2=95.4	0.284 0.839 j=2 r^2=96.6

APPENDIX I-A

ESTIMATION OF COEFFICIENTS

Complex coefficients $d_j = a_j + i\, b_j$, for $j = 0, 1, \ldots, N$, are to be estimated from a set of $\nu = N + 1$ complex numbers η_h so as to minimize the sum of the squared differences

$$S^2_{\eta;M} = \sum_{h=0}^{N} S^2_{\eta;h,M} = \sum_{h=0}^{N} (\eta_h - \eta_{h;M})\ (\eta_h - \eta_{h;M})^* \qquad \text{(A-1)}$$

for each index set M containing $1 \leq m \leq \nu$ elements, when the estimators $\eta_{h;M}$ are obtained from

$$\eta_{h;M} = \sum_{j \in M} d_j \exp\ (i\lambda\ jh), \qquad \lambda = 2\pi/\nu. \qquad \text{(A-2)}$$

The ν numbers $\{\eta_h\}$ are assumed to represent values or observations at ν equal intervals $h = 0, 1, \ldots, N$. These may be intervals of time or space; in the specific applications to be made here, they are equal intervals of height, and the numbers $\{\eta_h\}$ represent wind vectors at successive levels in the atmosphere. These vectors are expressed as departures from a plane of best fit, in the sense of minimizing variance, to the basic data; that is, any linear trend with height has been removed.

For each value of h

$$S^2_{\eta;h,M} = (\eta_h - \eta_{h;M})\ (\eta^*_h - \eta^*_{h;M})$$

$$= \left(\eta_h - \sum_M d_j\ e^{i\lambda jh}\right)\left(\eta^*_h - \sum_M d^*_j\ e^{-i\lambda\ jh}\right) \qquad \text{(A-3)}$$

$$= w^2_h + \left|\sum_M d_j\ e^{i\lambda\ jh}\right|^2 - 2\tilde{R}\left(\eta_h \sum_M d^*_j\ e^{-i\lambda\ jh}\right)$$

---

* The asterisk, *, denotes the complex conjugate.

because $\eta_h \eta_h^* = |\eta_h|^2 = w_h^2$, in the notation of Section 1. Since $\exp [i\lambda\, h(j - k)] = 1$ when $j = k$, the second term becomes

$$\left| \sum_{j \in M} d_j\, e^{i\lambda\, jh} \right|^2 = \sum_{j \in M} \sum_{k \in M} d_j\, d_k^*\, e^{i\lambda\, h(j-k)}$$

$$= \sum_{j \in M} |d_j|^2 + \sum_{j \neq k} \sum d_j\, d_k^*\, e^{i\lambda\, h(j-k)}. \qquad \text{(A-4)}$$

Expression of $\eta_h \exp (- i\lambda\, jh)$ as $\alpha_{hj} + i\beta_{hj}$ permits the final term in (A-3) to be written as

$$\eta_h \sum_{j \in M} d_j^*\, e^{-i\lambda\, jh} = \sum_{j \in M} d_j\, \eta_h\, e^{-i\lambda\, jh}$$

$$= \sum_{j \in M} (a_j - i\, b_j)\, (\alpha_{hj} + i\beta_{hj}). \qquad \text{(A-5)}$$

Since $|d_j|^2 = a_j^2 + b_j^2$ and $\Sigma \exp [i\lambda\, h(j - k)] = 0$, the sum of squares (A-1) to be minimized becomes

$$S^2_{\eta;M} = \sum_{h=0}^{N} S^2_{\eta;h,M} = \sum_{h=0}^{N} w_h^2 + \nu \sum_{j \in M} (a_j^2 + b_j^2)$$

$$- 2 \sum_{h=0} \sum_{j \in M} (a_j\, \alpha_{hj} + b_j\, \beta_{hj}). \qquad \text{(A-6)}$$

The usual minimization procedures give, for each value of j,

$$\frac{\partial S^2_{\eta;M}}{\partial a_j} = 2\nu\, a_j - 2 \sum_{h=0}^{N} \alpha_{hj},$$

$$\frac{S^2_{\eta;M}}{\partial b_j} = 2\nu\, b_j - 2 \sum_{h=0}^{N} \beta_{hj}. \tag{A-7}$$

Setting these derivatives equal to zero gives

$$a_j = \frac{1}{\nu} \sum_{h=0}^{N} \alpha_{hj}, \qquad b_j = \frac{1}{\nu} \sum_{h=0}^{N} \beta_{hj}. \tag{A-8}$$

Consequently,

$$d_j = \frac{1}{\nu} \sum_{h=0}^{N} (\alpha_{hj} + i\, \beta_{hj}) = \frac{1}{\nu} \sum_{h=0}^{N} \eta_h \exp\, (-\, i\lambda\, jh). \tag{A-9}$$

For computation, the real and imaginary parts are evaluated separately:

$$a_j = \frac{1}{\nu} \sum_{h=0}^{N} [u_h \cos\, (\lambda\, jh) + v_h \sin\, (\lambda\, jh)],$$

$$b_j = \frac{1}{\nu} \sum_{h=0}^{N} [v_h \cos\, (\lambda\, jh) - u_h \sin\, (\lambda\, jh)]. \tag{A-10}$$

In polar coordinates,

$$a_j = \frac{1}{\nu} \sum_{h=0}^{N} w_h \cos (\xi_h - \lambda\, jh),$$

$$b_j = \frac{1}{\nu} \sum_{h=0}^{N} w_h \sin (\xi_h - \lambda\, jh). \qquad \text{(A-11)}$$

Use in (A-2) of any set of m of these values for $d_j = a_j + i\, b_j$ will insure that the resulting estimator, $\eta_{h;M}$, when introduced into (A-1), will minimize the sum of squares $S^2_{\eta;M}$. When $m = \nu$, i.e., when the sum (polynomial) has as many terms as the original observations, $S^2_{\eta;M} = 0$. For smaller sets, i.e., for $m < \nu$, the sum of squares $S^2_{\eta;M}$ will depend on the exact composition of the set M. Thus, $S^2_{\eta;M}$ can be computed for each of the ν sets M in which $m = 1$, i.e., for one term only, and for the $\nu(\nu + 1)/2$ sets of two terms each, and so on, to find the combination giving an acceptably small $S^2_{\eta;M}$ from the smallest set M.

However, when the coefficients $\{d_j\}$ are orthogonal, in the statistical sense, the contribution of each is independent of that of the others, and

$$S^2_{\eta;M} = \sum_{j \in M} S^2_{\eta;j}. \qquad \text{(A-12)}$$

Then, $S^2_{\eta;j}$ can be computed for each orthogonal d_j and ranked in descending order to determine the minimum set M for which $S^2_{\eta;M}$ is acceptably small. The extent to which the coefficients $\{d_j\}$, estimated by (A-9), (A-10), or (A-11), satisfy these requirements is examined in Appendix B.

APPENDIX I-B

ORTHOGONALITY

Two different kinds of orthogonality are involved in the development of complex Fourier polynomials for the respresentation of wind profiles. One kind is that of the series of orthogonal polynomials used to represent a sounding. In such representation, functional orthogonality requires that

$$\sum e^{i\lambda\, jh}\, e^{i\lambda\, kh} = \begin{cases} \nu, & j = k \\ 0, & j \neq k. \end{cases}$$

Use of such a system of orthogonal functions permits judgement of the adequacy of the representation in terms of the sum of the squares of the coefficients. This sum measures the sum of the squares of the differences between the polynomial representation and the function being fitted, after removal of linear trend. When orthogonal functions are used, a smaller number of terms can be selected without recomputation of coefficients.

Another kind of orthogonality appears when a sounding is viewed as a collection of random variables. Then the coefficients $\{d_j\}$ in the Fourier representation (5.7) are also random variables, since they are linear combinations of the original random variables (6.2). Orthogonality of the system of coefficients $\{d_j\}$ is tantamount to their being uncorrelated. Uncorrelated Gaussian random variables are statistically independent - a very highly desirable property in computing probability statements. The basic physical quantities, i.e., balloon displacements or wind speeds, expressed in cartesian coordinates, are usually assumed to be approximately Gaussian. Hence the coefficients $\{d_j\}$, being linear combinations of them, also should be approximately Gaussian, especially because of central limit effects.

Orthogonality of the $\{d_j\}$ is almost impossible to establish unless the $\{\eta_h\}$ are second-order stationary with a real covariance function. The need for second-order stationarity, that is, that the covariance of (η_h, η_ℓ) depend only on the difference $|h - \ell|$, appears in the evaluation of the expression for the variances of the individual d_j. When the expectations of the $\{\eta_h\}$, and hence of the $\{d_j\}$, are zero, the variance of each d_j is given by

$$E(d_j\, d_j^*) = \nu^{-2} \sum_{h=0}^{N} \sum_{\ell=0}^{N} \exp\,[i\lambda\, j(\ell - h)]\, E(\eta_h\, \eta_\ell^*). \qquad \text{(B-1)}$$

* The asterisk, *, denotes the complex conjugate.

This involves the covariance of the observed values, or their departures from the plane, which in turn depends on the correlation (r) between the two components:

$$E(\eta_h\ \eta_\ell^*) = E\Big[(u_h u_\ell + v_h v_\ell) + i(u_\ell v_h - u_h v_\ell)\Big]$$
$$= \Big[r(u_h,u_\ell) + r(v_h,v_\ell)\Big] + i\Big[r(u_\ell,v_h) - r(u_h,v_\ell)\Big]. \qquad \text{(B-2)}$$

Second-order stationarity requires that these correlations depend, for each variable u or v, and for any separation h - ℓ, denoted as τ, only on the separation:

$$r(u_h,\ u_\ell) = r_u(h - \ell) = r_u(\tau) = r_u(-\tau). \qquad \text{(B-3)}$$

Certain properties of the separation τ are needed:

$$\tau = h\ -\ \ell, \qquad -N \leqq \tau \leqq +N,$$
$$\max(-\tau,\ 0) \leqq \ell \leqq \min(N - \tau,\ N). \qquad \text{(B-4)}$$

In this notation, (B-2) becomes

$$E(\eta_h\ \eta_\ell^*) = \Big[r_u(\tau) + r_v(\tau)\Big] + i\Big[r_{uv}(-\tau) - r_{uv}(\tau)\Big] = C(\tau), \qquad \text{(B-4)}$$

where C may be called a correlation function; C(0) = 2, because $r_u(0) = r_v(0) = 1$. In terms of this function C, the expression (B-1) for the variance is

$$E(d_j\ d_j^*) = \nu^{-2} \sum_{-N}^{+N} C(\tau)\ \exp\ (i\lambda\ j\tau)\ (\nu - \tau)$$
$$= \nu^{-1} \sum_{-N}^{+N} (1 - \tau/\nu)\ \cos\ (\lambda j\tau)\ \tilde{\tilde{R}}[C(\tau)] \qquad \text{(B-5)}$$
$$= \nu^{-1}\left[C(0) + 2 \sum_{1}^{N} (1 - \tau/\nu)\ \cos\ (\lambda j\tau)\ \tilde{\tilde{R}}\ C(\tau)\right]$$

because the variance is real-valued. [All summations are over the range of τ given by (B-4).] Similarly, the covariance function for the coefficients is

$$E(d_j\, d_k^*) = \nu^{-2} \sum_{h=0}^{N} \sum_{\ell=0}^{N} \exp\left[i\lambda\,(k\ell - jh)\right] E(\eta_h\, \eta_\ell^*)$$

$$= \nu^{-2} \sum_{\tau=-N}^{N} C(\tau) \exp\,(-\,i\lambda\, j\tau) \sum_{\ell} \exp\,(i\lambda\, \ell p), \qquad \text{(B-6)}$$

$$p = k-j,$$

because $k\ell - jh = \ell p - j\tau$. This must be zero for d_j and d_k to be orthogonal. To determine whether such is the case, (B-6) must be examined term by term, invoking the orthogonality properties of trigonometric series.

Since

$$\sum_{j=0}^{m} r^j = \frac{1 - r^{m+1}}{1 - r},$$

$$\sum_{j=0}^{N} \exp\,(i\lambda\, jp) = \begin{cases} \dfrac{1 - [\exp\,(i\lambda p)]^{\nu}}{1 - [\exp\,(i\lambda p)]}, & p \neq 0, \\ N + 1 = \nu, & p = 0. \end{cases} \qquad \text{(B-7)}$$

The last summation in (B-6), over ℓ, is, by definition (B-4), from max $(-\tau, 0)$ to min $(N - \tau, N)$, and hence depends on τ as well as on p. It may be denoted as $\gamma(\tau, p)$:

$$\gamma(\tau, p) = \sum_{\ell} \exp\,(i\lambda\ell p) = \begin{cases} 0 & \begin{cases} \tau = 0, \\ 0 \leq \ell \leq N; \end{cases} \\ \dfrac{1 - \exp\,(-i\lambda\, p\tau)}{1 - \exp\,(i\lambda p)} & \begin{cases} \tau > 0, \\ 0 \leq \ell \leq N - \tau; \end{cases} \\ -\dfrac{1 - \exp\,(-i\lambda\, p\tau)}{1 - \exp\,(i\lambda p)} & \begin{cases} \tau < 0, \\ -\tau \leq \ell \leq N. \end{cases} \end{cases} \qquad \text{(B-8)}$$

Thus, $\gamma(-\tau, p) = -\gamma(\tau, p)$. In the expression for $\gamma(\tau, p)$ when $\tau > 0$, multiplication of numerator and denominator by $1 - \exp(-i\lambda p)$ gives

$$\gamma(\tau, p) = \frac{(1 - e^{i\lambda p})(1 - e^{-i\lambda p\tau})}{2 - 2\cos \lambda p} \qquad \tau > 0. \tag{B-9}$$

This is zero when $p\tau$ is an integral multiple of ν and is small for τ such that $p\tau$ is close to an integral multiple of ν.

Next, the correlation function $C(\tau)$ must be examined. It is real if and only if it is even, i.e., if $r_{uv}(-\tau) = r_{uv}(\tau)$. In this case, (B-6) becomes

$$\begin{aligned} E(d_j d_k^*) &= \frac{1}{\nu}\left\{ \sum_{\tau=1}^{N} C(\tau)\, e^{-i\lambda\tau j}\, [\gamma(\tau,p)/\nu] + \sum_{\tau=1}^{N} C(-\tau)\, e^{i\lambda\tau j}[-\gamma(\tau,p)/\nu]\right\} \\ &= \frac{1}{\nu}\left\{- \sum_{\tau=1}^{N} C(\tau)\, [\gamma(\tau,p)/\nu]\, [e^{i\lambda\tau j} - e^{-i\lambda\tau j}]\right\} \\ &= \frac{1}{\nu}\left\{- \sum_{\tau=1}^{N} C(\tau)\, [\gamma(\tau,p)/\nu]\, [2i \sin j\lambda\tau]\right\} \\ &= 0\left(\frac{1}{\nu}\right). \end{aligned} \tag{B-10}$$

The summand of equation (B-10) will not be large since $|C(\tau)| \leq C(0) = 2$ and tends to zero as τ becomes large. The multiplier $\sin i\lambda\tau$ will have a dampening effect for the smaller values of τ.

Thus, $E(d_j d_k^*)$ apparently is always small, although that it is identically zero for all τ, as is required for complete orthogonality, has not been proved. Actually, $E(d_j d_k^*) \to 0$ as $N + 1 = \to \infty$, i.e., as more and more levels are used and the discrete model approaches a continuous one. Thus, the question of orthogonality may be analogous to the general problem of the extent to which large sample theory can be used for small samples, or to which properties of a continuous function can be applied to a discrete one. For the present purpose, the assumption of orthogonality seems reasonable.

MATHEMATICAL WIND PROFILES

PART II

SUMMARY

The application and properties of the augmented Fourier polynomial are examined in detail. Seasonal differences, differences between the velocity and momentum representations of the wind, length and specific interval in a sounding, and the effect of averaging and normalizing on a profile are treated in the discussion. Numerous tables are presented for comparative purposes. The basic data examined were two 34-kilometer soundings and two sequences of 6-hour soundings from Montgomery, Alabama.

1. INTRODUCTION (II)

Mathematical representation of the vertical profile of wind by augmented Fourier polynomials was proposed in Mathematical Wind Profiles , hereafter referred to as MWP-I. In that report, the basic problems of wind representation were discussed, and the augmented Fourier polynomials were developed. In such an exploratory study, many questions were posed which could not be answered immediately. Some of these questions are discussed further in the present (second) report, which also explores some problems mentioned in MWP-I as requiring additional investigation.

1.1 Model

For completeness, the basic properties of augmented Fourier polynomials are summarized here from MWP-I. The eastward (x) and northward (y) components of the horizontal wind at a given level are combined into a complex variable $z = x + iy$. A new complex variable $w = u + iv$, obtained from z by the removal of linear trends, utilizes the residuals u and v, which have zero means. The Fourier representation of this new variable w is

$$w(h) = \sum_{j=0}^{n} d_j \exp(i\lambda j h), \qquad h = 0, 1, \ldots, n, \tag{1.1}$$

where $\lambda = 2\pi/\nu$ and $\nu = n + 1$. Each complex Fourier coefficient d_j is composed of a real and imaginary part: $d_j = a_j + ib_j$. The d_j are estimated by the method of least squares as

$$\begin{aligned} a_j &= \frac{1}{\nu} \sum_{h=0}^{n} \left[u_h \cos(\lambda j h) + v_h \sin(\lambda j h) \right] \\ b_j &= \frac{1}{\nu} \sum_{h=0}^{n} \left[v_h \cos(\lambda j h) - u_h \sin(\lambda j h) \right] . \end{aligned} \tag{1.2}$$

For $j = 0$, $a_o = b_o = 0$ because $\bar{u} = \bar{v} = 0$. The variance of w is given by

$$s_w^2 = \frac{1}{\nu} \left[\sum_{h=0}^{n} u_h^2 + \sum_{h=0}^{n} v_h^2 \right] = \sum_{j=0}^{n} A_j^2 , \tag{1.3}$$

where $A_j^2 = a_j^2 + b_j^2$. The variance of the original variable z is computed as

$$s_z^2 = s_x^2 + s_y^2 \tag{1.4}$$

where $s_x^2 = \nu^{-1} \sum_{h=0}^{n} (x_h - \bar{x})^2$ and $s_y^2 = \nu^{-1} \sum_{h=0}^{n} (y_h - \bar{y})^2$.

The percent reduction in variance by removal of trend lines in x and y, referred to as percent variance explained by the linear term, is

$$100\, s_L^2 = 100\,(1 - s_w^2/s_z^2). \tag{1.5}$$

The percent of the total variance explained by the jth harmonic is

$$100\, s_j^2 = 100\, A_j^2/s_z^2 . \tag{1.6}$$

The partitioning of the total variance of w (1.3) into ν parts in terms of the complex Fourier coefficients is a consequence of the Parseval identity. Orthogonality of the d_j's implies independence and allows for meaningful partitioning of the total variance into components of variance for each harmonic. The d_j's are shown to be almost orthogonal in Appendix B of MWP-I. The percent variance explained by each harmonic is a measure of that harmonic's importance in the mathematical description of the profile.

1.2 Harmonics

The method was applied in MWP-I to twelve monthly averages from Cape Kennedy and to four soundings at consecutive six-hour intervals in January, 1956 from a two-year (1956-1957), winter (Dec., Jan., Feb.) and summer (June, July, Aug.) set of data from Montgomery, Alabama. In the Montgomery data, both velocity and momentum representations were used. All cases were analyzed in terms of 24-point profiles of 2 through 25 km. The total variance in each case was composed of contributions from the linear term and 23 harmonics. In these 20 cases analyzed in MWP-I, only the linear term plus four selected harmonics were needed to explain 85 to 95 percent of the total variance.

The original intention was to base the final regression equations for prediction on a reduced number of selected harmonics. However, subsequent computer runs have shown that only about two seconds of IBM 7094 time are required per profile for computation of all quantities needed in complex Fourier analysis. This includes the trend lines in x and y and their removal, all complex Fourier coefficients, and the percent variance explained by the linear term and each harmonic. Use of all harmonics for regression thus seems more advantageous because of the perfect fit thereby given, with negligible increase in computation time.

However, is some cases, the use of all coefficients, linear trend and complex Fourier, may not be possible. The number of profiles available for

estimating regression coefficients may be less than the total number of coefficients; computer capacity may be inadequate to solve sets of linear equations of large order. Linear regression for the complete representation requires estimation of regression coefficients which number twice the number of data points plus three. A minimum of N observations is required to estimate N regression coefficients, but for good regression estimates the number of observation should be much larger than the number of coefficients to be estimated. In many cases, sufficient observations may not be available, and the reduced harmonic representation should be used.

The major computer operation in linear regression is the solution of a linear set of equations, say p equations in p unknowns, which in matrix form requires the capacity to invert a matrix of order p. Inversion of a normal matrix of order 100 can be handled in core on the IBM 7094. Although a matrix of order 450 to 500 can be inverted with considerable time and difficulty, the small magnitude of some Fourier coefficients, plus programming difficulty, suggest about 100 as the maximum number of regression coefficients that could be used for prediction. While the present examination of 24-point profiles can be handled easily, soundings with more than 48 points would cause problems. Thus, for representation and prediction of a long enough sounding with data spacing of less than 1/2 km a reduced number of harmonics is almost mandatory.

Although all harmonics may ultimately be used for the present study, sufficient information exists to point out the importance of studying various features of the reduced harmonic representation. The relative importance of the various harmonics and their contribution to the variance were studied for additional soundings to supplement the results of MWP-I and to provide information about the method of reduced harmonics. The discussion, in terms of what was found in particular representations, may provide sufficient guides to more important developments.

1.3 Applications

In this further application and extension of the previous work, the complex Fourier method is used to represent and examine the two highest winter and summer soundings of the two-year Montgomery, Alabama data. These representations involve 11 overlapping 24-point partitions of each soundings. Two additional series of 6-hour observations over periods of 24 hours or longer and to 25 km or higher also are studied.

The basic wind data for each sounding were represented and analyzed separately in terms of velocity and momentum. Velocity is represented by the x and y Cartesian components of the wind at consecutive heights. The momentum representation is the product of the velocity (x, y) and the atmospheric density ρ at sequential heights and has units of dynes per cubic centimeter. The importance of momentum as a predicting variable has not as yet been established.

Often, over certain intervals in the atmosphere, an increase in wind speed with height is accompanied by a proportionate decrease in atmospheric density with height, which implies that the momentum is constant throughout the interval. This situation, observed many years ago, forms the basis for "Egnell's law" (see MWP-I) which states that between 5 and 10 km the momentum is constant. Therefore, momentum may be a better predicting variable than velocity, provided it is constant over a large enough portion of the prediction interval. In MWP-I the momentum representations of the 6-hour Montgomery data showed larger cumulative percent variance explained than the velocity representations when compared for an equal number of harmonics.

2. SCOPE OF WORK AND BASIC DATA

The material presented in this section is primarily descriptive. Empirical results, in the form of tables, are used to explore some questions posed in MWP-I and to supplement the work of that report. Basic considerations, in addition to application of the method, are: seasonal differences, velocity-momentum differences, importance of the first 2 km, importance of various 24-km intervals in a longer profile, differences in importance between arbitrarily selected harmonics and those observed, and the length of the profile in variance explanation. By necessity most of the results are very general and may have serious limitations because of the small sample size.

Complex Fourier analysis was performed on the highest winter and summer soundings in the two years of Montgomery data on hand. These were for 0300Z on 12 February and 1000Z on 8 June, 1956 both from 0 through 33 km. The basic data for both are given in Appendix A. In all, 52 Fourier representations were computed from the two soundings. These were for 0 through 33 km, 2 through 33 km, and 0 through 23 km, 1 through 24 km, ..., 10 through 33 km, on each sounding separately for velocity and momentum.

2.1 Explained Variance

In MWP-I the large amount of cumulative percent variance explained, hereafter referred to as CPVE, by a few harmonics was of particular interest in the representation and analysis of the monthly wind averages from Cape Kennedy and the 24-hour sequence in January from Montgomery. Differences in CPVE were observed between the velocity and momentum representations of the wind in both sets of data. Monthly comparisons of the Cape Kennedy data showed some seasonal effect on the CPVE; seasonal comparisons were not possible for the routine 6-hour soundings for Montgomery because only one sequence was computed for January. From Eqs. (1.5) and (1.6)

$$(\mathrm{CPVE})_{(J)} = 100 \left(s_L^2 + \sum_{j=1}^{(J)} s_j^2 \right) . \qquad (2.1)$$

The preliminary nature of MWP-I left unanswered the question whether the observed differences were real, or were due to spurious sampling fluctuation. The answer to this question would provide insight into the method of representing the wind and the possible seasonal limits to prediction.

Another question of importance left unanswered was that of how the CPVE changes with the particular interval and length of interval chosen. Whether one particular subset of a wind profile has a greater CPVE in a limited Fourier representation than any other was not determined in MWP-I. The amount of CPVE apparently varies inversely with the length of the interval when a reduced number of harmonics is used for the representation; this will be discussed later in this report.

Of special interest in the interval representation are the surface and first kilometer. These were excluded in MWP-I because of assumed friction effects, and representation began at the 2-km level.

First, the decomposition of each of the 34-point soundings (momentum and velocity for each season) into eleven profiles of 24 points each is considered. Information about the representation of the total profiles is given later in this report. Table 1 gives the CPVE by the linear term alone and the linear term plus the four or six highest ranked harmonics in variance explanation. In a 24-point profile with zero mean 23 harmonics are possible; thus, four and six terms represent 17.4% and 26.1% of the possible number of harmonics.

The results shown in Table 1 are only indications of a more complicated situation and are applicable only for the special case of a reduced number of harmonics based on a particular sample of two soundings. The results do have additional value as examples of the application of the complex Fourier method of representing a wind vector profile. Only the CPVE by the linear plus four highest ranked harmonic terms will be discussed. The linear plus six highest ranked harmonic terms gives approximately the same indications about seasonal and velocity-momentum relationships, although the inclusion of two extra terms gives a larger CPVE and a smaller range of CPVE values, as expected. The particular height interval used in the discussion will be designated by its lower and upper limits, with a dash between, e.g., 2 through 25 km will be written 2-25.

TABLE 1

Cumulative Percent Variance Explained (CPVE) by Linear, Linear Plus Four, and Linear Plus Six Highest Harmonics, Respectively, for Two Soundings at Montgomery, Alabama

PROFILE INTERVAL (km)	WINTER (12 Feb 56)						SUMMER (8 Jun 56)					
	Velocity			Momentum			Velocity			Momentum		
	L	L+4	L+6	L	L+4	L+6	L	L+4	L+6	L	L+4	L+6
0-23	4	95	96	32	89	93	4	92	95	16	77	83
1-24	10	95	96	51	89	92	4	91	94	31	85	90
2-25	21	89	91	67	89	92	7	90	92	35	86	91
3-26	35	87	91	80	93	94	12	89	92	38	86	91
4-27	49	88	91	86	95	96	15	84	90	40	82	89
5-28	61	89	92	86	96	97	19	80	88	47	79	87
6-29	69	90	92	85	97	98	29	83	88	65	85	89
7-30	70	89	92	82	98	98	38	87	91	72	89	92
8-31	65	87	90	77	98	99	43	88	91	76	95	96
9-32	64	89	92	73	98	99	41	84	89	73	97	97
10-33	59	92	94	68	98	99	33	85	92	68	97	98

In general the CPVE values in Table 1 show a considerable degree of variation with interval. The 12 Feb 56 momentum values increase monotonically with increasing height interval; the other three soundings show no consistent increase or decrease with interval height. A more complicated analysis of the pattern of variation of CPVE with interval is not warranted; this depends critically on the number and importance of the frequency components of the original wind. This decomposition into frequency components is discussed later in terms of similarity of the harmonic numbers observed.

Table 2 presents the maximum, minimum, and range of CPVE values found in Table 1.

TABLE 2

Maximum, Minimum, and Range of CPVE from Table 1

		Max. (%)	Min. (%)	Range (%)	Range (as % of Max.)
WINTER	Velocity	95	87	8	8.4
	Momentum	98	89	9	9.2
SUMMER	Velocity	92	80	12	13.0
	Momentum	97	77	20	20.6

The range of CPVE for velocity is less than that for momentum, while the winter ranges for both are considerably less than the summer. Momentum, although having a larger range, has a larger maximum CPVE value than velocity in both seasons.

In winter, the maximum CPVE for velocity is found in the 0-23 and 1-24 intervals, while the minimum CPVE for momentum is found in the 0-23, 1-24 and 2-25 intervals. In summer the maximum CPVE for velocity and minimum CPVE for momentum occur in the 0-23 intervals for winter and the 5-28 intervals for summer; a secondary CPVE minimum for momentum is also found in this same 5-28 interval. CPVE maxima for momentum are found in intervals 7-30,...., 10-33 for winter and 9-32 and 10-33 for summer.

Four meaningful comparisons can be made by using the differences of CPVE at each level among velocity, momentum, and season. In the winter at the 3-26 interval and above, CPVE for momentum is larger than that for velocity, where a maximum difference of 11% is obtained in the 8-31 interval.

In summer CPVE for velocity is larger from 0-23 through 5-28 with a maximum difference of 15% at 0-23; CPVE for momentum is larger from 6-29 through 10-33 with a maximum difference at 9-32 of 13%.

The comparison of CPVE for velocity in each interval between winter and summer shows, excepting for three intervals where the difference is small, that larger values of CPVE occur in the winter sounding. CPVE for momentum in winter is larger than that for summer in all intervals and considerably larger in all except the 9-32 and 10-33 intervals.

If a reduced number of harmonics is to be used, indications based on this limited sample are that CPVE, and hence predictability, can vary considerably with the interval chosen. While in practice the interval probably would be chosen on the basis of physical importance, the information given here can provide guidelines for a choice among several alternatives. For example, velocity shows more CPVE in the intervals containing the first two kilometers, while momentum CPVE increases away from intervals containing the first two kilometers. In terms of CPVE, indications for season and type of representation are that winter will be better than summer for both velocity and momentum, and in both seasons momentum will be better than velocity, excluding the near surface layers.

2.2 Ranking Order

The CPVE discussed in the previous section is important for indications about the limits of predictability in terms of season and type of representation. This section deals with the ranked harmonic numbers associated with these CPVE values. If a reduced number of harmonics is to be used for prediction, some indication is needed as to the relative importance of each harmonic in terms of season and the velocity and momentum representations and how they change with the interval chosen. This would provide clues to the specific harmonic numbers to be used in a predicting equation. Harmonic numbers are given in Table 3 for the six highest ranked harmonics, in terms of percent variance explained both for velocity and momentum, for the eleven 24-point profiles of the 12 Feb 56 and 8 Jun 56 soundings.

In the winter (12 Feb 56) sounding, the harmonic numbers for velocity and momentum are almost identical in the first two ranks. Only the 1st harmonic is found in the first rank, while in second rank the 23rd dominates along with a few 22nd harmonics. The third and fourth ranks are mixed, with the 2nd and 22nd harmonics being the most frequently present. The fifth and sixth ranks are completely mixed.

In the summer (8 Jun 56) sounding, the first three ranks of velocity and momentum are dominated by the 1st, 2nd, and 23rd harmonics, with the 2nd harmonic replacing the 1st as most important in the middle intervals for momentum. The fourth rank is mixed but is dominated by the 21st

TABLE 3

First Six Ranked Harmonic Numbers, on Basis of Percent Variance Explained

PROFILE	WINTER 12 Feb 56						SUMMER 8 Jun 56					
	RANKS											
INTERVAL (km)	I	II	III	IV	V	VI	I	II	III	IV	V	VI
						Velocity						
0-23	1	23	15	16	10	4	1	23	2	20	21	4
1-24	1	23	15	14	9	7	1	23	2	21	20	7
2-25	1	23	16	15	6	11	1	23	2	21	19	20
3-26	1	23	2	22	3	16	1	23	2	21	20	4
4-27	1	23	2	22	3	15	1	2	23	21	20	22
5-28	1	22	23	2	3	21	1	2	23	21	22	20
6-29	1	22	2	23	6	16	1	2	22	4	21	23
7-30	1	22	2	15	18	5	2	1	22	3	4	19
8-31	1	23	22	19	2	5	1	2	22	17	3	7
9-32	1	23	3	19	22	15	1	2	23	20	21	22
10-33	1	23	3	15	14	4	1	23	2	21	20	4
						Momentum						
0-23	1	23	2	4	3	22	23	1	2	4	3	6
1-24	1	23	22	2	4	21	1	23	2	22	21	7
2-25	1	23	22	2	21	5	1	23	2	22	21	7
3-26	1	22	2	16	21	3	2	1	23	21	22	20
4-27	1	22	2	3	21	9	2	1	23	21	22	20
5-28	1	23	22	2	17	16	2	1	23	21	20	22
6-29	1	23	2	4	22	16	2	23	1	20	3	4
7-30	1	23	9	15	4	2	2	1	23	5	3	6
8-31	1	23	16	17	2	10	1	23	2	17	7	3
9-32	1	23	2	15	22	16	1	23	2	21	20	16
10-33	1	23	22	2	16	15	1	23	2	19	17	20

harmonic. The fifth and sixth rankings are mixed, the 21st, 20th, and 22nd harmonics being most frequent.

In both soundings, the harmonic numbers in the first four ranks are consistent over all intervals for each of the four situations. As the specific interval changes, the order of the harmonics may change slightly. The fifth and sixth ranks do not seem to be consistent over all intervals but are usually represented by harmonics which contribute little to the CPVE. This indicates that an appropriate choice of two extra harmonics for prediction would be difficult. The similarity of all 23 ranked harmonics from interval to interval will be discussed later.

Overall, the harmonic numbers shown in Table 3 are quite similar to those found in MWP-I, where the ranking order was usually the 1st, 23rd, 2nd, and 22nd harmonics. Considering velocity and momentum for each sounding separately, the order and importance of the harmonic numbers change slightly in this limited sample. A compilation of the frequency of occurrence of harmonic numbers for velocity and momentum with season is given in Table 4. Only the four highest frequencies of occurrence have been tabulated. A harmonic number could appear a maximum of 11 times.

TABLE 4

Frequency of Occurrence of the Four Most Important Harmonic Numbers

Harmonic Number		1	23	2	22	21	15
12 Feb 56	Velocity	11	10	5	6	-	5
	Momentum	11	9	9	6	-	-
8 Jun 56	Velocity	11	8	11	-	6	-
	Momentum	11	11	11	-	4	-

The actual rank order of the harmonics for prediction is not important unless the Fourier coefficients of each are to be weighted in some special way. Therefore, the three most important harmonics for this sample would be 1, 23, and 2, with 22 for winter and 21 for summer as fourth most important.

2.3 Effect of Arbitrary Selection

In MWP-I, certain harmonic numbers seemed to be consistently important in CPVE. The difference in CPVE which would result if pre-

selected harmonics were used rather than the actual ranked harmonics was studied for this report. Based on MWP-I, the first four preselected harmonics used were the 1st, 23rd, 2nd, and 22nd, with the 3rd and 21st added for the fifth and sixth terms. The difference in CPVE by the best actual and by the preselected harmonics is given in Table 5, with the median and average over all 11 intervals. The values are given as percent of total variance explained, and are the amount that the CPVE by the preselected terms was less than that by the actually best terms.

TABLE 5

Difference in CPVE by Preselected and by Actually Best Harmonics, Given as Percent of Total Variance Explained for 4 and 6 Harmonic Terms

PROFILE INTERVAL (km)	12 Feb 56				8 Jun 56			
	Velocity		Momentum		Velocity		Momentum	
	4	6	4	6	4	6	4	6
0-33	1.5	2.5	0.8	2.1	1.8	2.9	1.0	1.9
1-24	1.9	2.4	0.0	0.3	2.1	2.1	0.0	0.6
2-25	0.8	1.3	0.0	0.1	3.0	1.6	0.0	0.8
3-26	0.0	0.6	0.2	0.2	3.7	1.5	0.3	0.1
4-27	0.0	0.5	0.4	0.2	3.3	2.7	0.7	0.8
5-28	0.0	0.0	0.0	0.1	1.0	2.8	1.1	2.5
6-29	0.0	0.4	0.1	0.4	1.2	1.3	1.3	1.7
7-30	0.9	3.1	0.2	0.4	2.2	2.7	1.7	3.0
8-31	0.8	2.2	0.3	0.6	1.3	2.1	0.5	1.0
9-32	0.7	0.9	0.0	0.3	0.4	1.9	0.4	0.4
10-33	2.2	1.2	0.0	0.4	5.5	6.1	0.5	1.0
Median	0.8	1.2	0.1	0.3	2.1	2.1	0.5	1.0
Average	0.8	1.4	0.2	0.5	2.3	2.5	0.7	1.3

The differences shown in Table 5 are surprisingly small. Except for the velocity in June, where the average difference is about 2% of the total variance, the average difference is less than 1% for four terms and about 1% for six terms for the other three stratifications. Use of preselected harmonics rather than computing all harmonics and then selecting a reduced number for prediction may be useful if large computers are not available.

2.4 Long Profiles

In this section the results for the total 0-33 km profile and the 2-33 km profile are considered for the February and June soundings for both velocity and momentum. Table 6 gives the CPVE by the linear, and the linear plus four, plus six, plus eight, and plus ten highest ranked variance-reducing harmonics.

TABLE 6

CPVE for Two Long Profiles, for Velocity and Momentum

PROFILE INTERVAL (km)			L	L+4	L+6	L+8	L+10
0-33	12 Feb 56	Velocity	15	84	88	91	93
		Momentum	50	84	89	92	94
	8 Jun 56	Velocity	19	80	86	90	93
		Momentum	37	75	80	82	84
2-33	12 Feb 56	Velocity	30	82	86	89	92
		Momentum	69	90	92	93	95
	8 Jun 56	Velocity	20	80	88	90	92
		Momentum	48	85	89	92	94

TABLE 7

Ranked Harmonic Numbers for Velocity and Momentum
(The first four profiles have a total of 33 harmonic numbers, the last four 31).

PROFILE INTERVAL (km)			RANKS I	II	III	IV	V	VI
0-33	12 Feb 56	Velocity	1	2	32	33	4	5
		Momentum	1	2	32	3	4	33
	8 Jun 56	Velocity	2	1	33	32	30	3
		Momentum	2	33	32	1	3	28
2-33	12 Feb 56	Velocity	1	30	2	31	4	6
		Momentum	1	2	30	29	3	6
	8 Jun 56	Velocity	2	1	31	30	28	27
		Momentum	2	31	30	3	1	28

The first six ranked harmonic numbers are shown in Table 7.

In general, indications in Table 6 and 7 are similar to those found in the earlier discussions of the 24-point profiles. The harmonic numbers follow about the same pattern: the first and last two harmonics dominate the CPVE. For the 0-33 km profile the CPVE for velocity in both February and June is about the same as the February momentum, while that for momentum in June is less. For the 2-33 km profile CPVE for velocity is about the same for both dates and less than for momentum, for which the CPVE is larger in winter than in summer. The influence of the first two kilometers is again shown: CPVE for momentum in the 2-33 km profile is larger than for the 0-33 km profile but for velocity is smaller.

For equal numbers of terms, the CPVE in these 32 and 34-point profiles is smaller than for the 24-point profiles: as the interval length increases the likelihood that the wind vector can be represented adequately by a few frequency components decreases. Thus, in the longer profile more frequency components are needed to describe mathematically the wind profile, and less CPVE is found in comparison with a shorter profile, which would require fewer components for representation.

3. SIMILARITY OF COMPLETE INTERVAL REPRESENTATIONS

The previous sections have been devoted to descriptive material about the February and June soundings and their decomposition into 24-point profiles. The discussion has been in terms of cumulative percent variance explained (CPVE) by various harmonic numbers. Another important aspect of the general problem is whether the complete Fourier representation changes over various portions of a long sounding. This question is discussed in detail in this section.

One problem of ranking in this case is that the first four or six ranked harmonics are large and differ considerably, while the rest are small and differ slightly. The question still remains as to whether the slight differences observed are real or are due mainly to sampling fluctuations.

Comparison of the complex Fourier coefficients for two overlapping intervals in the same sounding is much more complicated statistically than the comparison of a similar representation of two different soundings. The major difficulty is that overlapping values cause the comparison to be made between profiles which are not independent. For example, when a 0-33 km sounding is decomposed into eleven 24-point profiles, adjacent intervals differ only in the higher value of each profile, and all eleven intervals contain values for 10 through 23 km. Effectively, the problem is one of examining the results of a moving average on a mathematical representation. Because of this moving average, the sum of the variance about x and y, and about the reduced zero-mean variables u and v, changes from interval to interval.

The complete set of complex Fourier coefficients for each of the eleven 24-point profiles obtainable from the two 33-km soundings was used for comparison. These were also compared for only the nine upper intervals, which excluded the 0-23 and 1-24 km profiles. A non-parametric method was used to eliminate some of the statistical problems of comparison and as a time saving device in computation and development of new methodology.

Each of the 23 possible harmonics in a zero-mean 24-point profile can be considered to be a rankable attribute in terms of the importance of the frequency it represents. Analogous in terms of parametrics would be a discrete vector periodogram of the profile, each point being a measure of that component of variance. The ranking of all the harmonics in a profile was on the basis of the size of the Fourier coefficients, which in turn is related to the percent variance explained by each harmonic. Ranking considers relative rather than absolute differences in size.

One of the oldest and perhaps the best known rank statistics is the Spearman rank correlation coefficient,

$$r_s = 1 - \frac{6}{N(N^2 - 1)} \sum_{j=1}^{N} \Delta_j^2 \qquad (3.1)$$

where Δ_j is the difference in ranks of the jth terms and N is the number of terms ranked. For comparison of two 24-point intervals, N is 23, and

$$r_s = 1 - \sum \Delta_j^2 / 2024$$

The probability of occurrence of any particular r_s value is the ratio of the number of permutations of rankings giving rise to that value to the N! rankings possible.

The Spearman rank correlation coefficient measures the agreement between two sets of rankings, but for this study the amount of agreement between eleven sets of rankings is of greater interest. One method of establishing the degree of relationship between multiple rankings would be to compute all possible rank correlations and then average them. For eleven intervals this is the combination of eleven things taken two at a time, or fifty-five coefficients to compute. Fortunately, another statistic, developed for comparison of multiple rankings, is less tedious to compute and is linearly related to the average of all rank correlations. The Kendall coefficient of concordance, W, is an index of divergence between the actual agreement shown in the data and the maximum or perfect agreement possible:

$$W = \frac{12S}{k^2 N(N^2 - 1)} \quad , \qquad (3.2)$$

where k is the number of ranked sets, N is the number of terms ranked, and S is a measure of dispersion among k sets of rankings given by

$$S = \sum_{j=1}^{N} \left(R_j - \frac{1}{N} \sum_{j=1}^{N} R_j \right)^2$$

$R_j = \sum_{m=1}^{k} R_{jm}$, where R_{jm} = the rank of the jth item in the mth set. The average of the Spearman rank correlation coefficients, denoted r_{sav}, has a range of values $-1 \leq r_{sav} \leq 1$ and is related to W, with range of $0 \leq W \leq 1$, by

$$r_{sav} = \frac{kW - 1}{k - 1} \quad . \tag{3.3}$$

The probability of occurrence of W for small N is found by enumerating all possible outcomes and then taking the ratio of the favorable to the total outcomes possible, as in the Spearman. For $N > 7$ the following statistic,

$$\frac{12S}{kN(N+1)} \simeq \chi^2_{(N-1)} \quad , \tag{3.4}$$

i.e. a chi-square statistic with N - 1 degrees of freedom.

The 23 harmonics in each of the 11 intervals (0-23 to 10-33) for velocity and momentum in both soundings were ranked on the basis of the size of the percent variance explained by each. Two W values were computed for each of the four decompositions, one for the 0-23 through 10-33 intervals and the other for 2-25 through 10-33 intervals. Two sets were used to determine the effect of the first two kilometers on the overall agreement of all interval representations.

The W and r_{sav} values, computed using equations 3.2 and 3.3, are given in Table 8.

TABLE 8

W and r_{sav} Values

		Intervals 0-23 through 10-33		Intervals 2-25 through 10-33	
		W	r_{sav}	W	r_{sav}
12 Feb 56	Velocity	.45	.40	.54	.49
	Momentum	.51	.46	.51	.44
8 Jun 56	Velocity	.65	.62	.66	.62
	Momentum	.60	.56	.60	.56

Tested by the statistic of equation 3.4, all W-values were found to significant at less than the 0.001 level, i.e., the probability that the observed W values could occur by chance is less than 1 in 1000.

For both velocity and momentum on the June sounding, W is higher than for the February sounding. Inclusion of the 0-23 and 1-24 intervals seems to make little difference in the amount of agreement found, with the possible exception of velocity in February where the 2-23 through 10-33 shows a larger W.

Although values of W are not close to 1.0, showing a perfect agreement, they are significant and high enough to indicate a large amount of consistency in the relative importance of the Fourier coefficients in all intervals.

Some of the deviation from perfect agreement arises because many Fourier coefficients in each interval are small and differ only slightly in magnitude. In the non-parametric ranking process, small differences are given the same weight as large differences. If these small differences could be established as due to sampling fluctuation, the small Fourier coefficients could be ranked as ties, which would increase the value of W.

In general, these limited observations indicate that the Fourier representation does not change markedly with the interval chosen in a long profile.

In the previous discussion, inclusion of the intervals containing the surface and first kilometer made no appreciable change in W. However, dissimilarity of the rankings for the 0-23 and 2-25 km intervals appeared in certain cases. As a check on the agreement between the rankings for the 0-23 and 2-25 intervals with the other ten intervals, Spearman rank correlation coefficients, denoted r_s, were computed for these comparisons (Table 9). For the 23 ranked harmonics, r_s values of 0.35 and 0.50 are significant at the 0.05 and 0.01 levels, respectively, for the one-tailed test of a positive correlation.

TABLE 9

Rank Correlation Between (0) and (2) Intervals with All Other Intervals

(Intervals are numbered 0 through 10. Velocity is denoted V and momentum M. The comparisons of the 0 through 23 km and 2 through 25 km intervals with the other intervals are denoted (0) and (2).)

	12 Feb 56				8 Jun 56			
INTERVAL	V(0)	V(2)	M(0)	M(2)	V(0)	V(2)	M(0)	M(2)
0	1.00	.26	1.00	.67**	1.00	.78**	1.00	.65**
1	.55**	.26	.55**	.70**	.76**	.92**	.65**	.96**
2	.26	1.00	.67**	1.00	.78**	1.00	.65**	1.00
3	.23	.49*	.62**	.55**	.74**	.85**	.71**	.89**
4	.15	.49*	.41*	.57*	.69**	.72**	.68**	.88**
5	.30	.32	.49*	.71**	.64**	.76**	.59**	.89**
6	.12	.56**	.77**	.70**	.54**	.62**	.72**	.49**
7	.06	.33	.23	.52**	.39*	.62**	.53**	.46*
8	.41*	.31	.15	.36*	.47*	.47*	.43*	.54**
9	.32	.40*	.47*	.36*	.30	.47*	.34	.34
10	.49*	.29	.45*	.52**	.80**	.88**	.26	.43*

* Significant at 0.05 Level

** Significant at 0.01 Level

In general, the r_s values given in Table 9 show that the Fourier representation for the 2-25 km interval is better related to the other interval representations than is the 0-23 km interval representation. In most cases the r_s values are slightly larger for the 2-25 km interval. The table also shows how one interval is related to another and how this relation changes with interval position in the longer soundings.

Thus, a more representative description of the profile may be obtained if the surface and first kilometer were not used in the wind sounding to be represented. Whether the second, third, etc. kilometers should also be eliminated for better representation could only be determined by more extensive analysis. The 2-km lower limit, however, seems reasonable on the basis of observed wind variability, theory, and practical considerations.

Although the representations over all intervals are related and similar, the physical uses of the prediction are more basic and important. What heights are critical? How does prediction reliability change in relation to the position of critical heights within the interval used for representation? Should all height information be used, or should the interval of representation be shortened? These questions will have to be answered when the final method of prediction is developed by studying the variability of predicted values with changing intervals.

4. SEQUENTIAL SIX-HOUR SOUNDINGS

In MWP-I results of complex Fourier analysis on four consecutive soundings at 6-hour intervals, from 0900Z 9 Jan 56 through 0300Z the next day at Montgomery, were given as an example of the method. This example has been supplemented by analysis of the only other two series of consecutive soundings from the Montgomery data with duration of 24 hours or longer and a maximum height, without gaps, of at least 25 km. These were for consecutive 6-hour intervals from 0300Z on 29 Jan 56 through 0300Z the next day and from 1200Z on 1 July 57 through 1800Z on the next day. The CPVE by the linear, linear plus four, and linear plus six ranked harmonics is given in Table 10.

TABLE 10

CPVE for Consecutive Six-Hour Soundings

Date	Hour	Velocity			Momentum		
		L	L+4	L+6	L	L+4	L+6
29 Jan 56	0300	2	92	95	47	89	91
29 Jan 56	0900	3	80	84	50	85	90
29 Jan 56	1500	16	69	76	64	84	88
29 Jan 56	2100	12	81	88	74	91	94
30 Jan 56	0300	16	74	83	74	89	92
1 Jul 57	1200	71	89	93	65	89	92
1 Jul 57	1800	75	88	91	61	87	91
2 Jul 57	0000	69	85	91	64	90	93
2 Jul 57	0600	60	84	88	71	91	94
2 Jul 57	1200	65	93	96	78	92	95
2 Jul 57	1800	78	91	93	76	91	93

Some aspects of Table 10 are a little surprising, while others are consistent with indications from MWP-I. As usual for a fixed number of terms, CPVE for momentum is greater than for velocity. However, CPVE for velocity and momentum for the summer values are larger than those for winter in this report. This is the opposite of the situation in MWP-I. The summer CPVE values of this report are smaller than those found for 9 Jan 56 in the first report, which had a range over the sequence for the linear plus four terms of 88 to 94% for momentum. For the 29 Jan 56 sequence, CPVE for the linear plus four terms ranges are 69 to 92% for velocity, and 84 to 93% for momentum, while the 1 July 57 sequence had 84 to 91% for velocity and 87 to 92% for momentum.

Thus, in a sequence of 6-hour soundings, the amount of CPVE can vary considerably. Indications of monthly and yearly variability are present. Perhaps the sequence analyzed in MWP-I was unusual, with a consistent pattern of harmonic numbers and large CPVE.

The ranked harmonic numbers associated with the CPVE in Table 10 are about the same as those found for the 9 Jan 56 sequence in MWP-I. The first and second most important harmonics are the 1st and 23rd. However, in the two new sequences, the 23rd was generally more important than the 1st, just the opposite of the 9 Jan 56 sequence. The third and fourth most important harmonics were mixed among the 2nd, 22nd, and 21st harmonic numbers. In general the harmonic numbers are not so consistent in order and frequency of appearance as those found in MWP-I.

One further analysis was performed on the sequential 6-hour soundings to examine how the Fourier representations vary with time. As a crude method of examining the consistency of the Fourier representation, Spearman rank correlations were computed for the ranked Fourier coefficients of all possible pairs in each sequence. The ultimate result of these r_s values was to be some indication of how the representation changed with selected time lags, which would also give some indication of best predicting times. With the short lengths of sequences available, the time lags would have to be limited to 6 and 12-hour lags.

Results were highly inconclusive, and the actual r_s values are not presented. The 9 Jan 56 profile from MWP-I showed good agreement for both the 6 and 12-hour lags for all times in the sequence for velocity, while the 6-hour lag was good but the 12-hour poor for momentum. In contrast, 29 Jan 56 and 1 Jul 57 sequences have highly variable r_s values. In general the r_s values were small even for some 6-hour lags. Persistence is being examined in more detail in another phase.

5. MODIFICATION OF OBSERVATIONS

Two questions about modification of the original observations were raised in MWP-I. The first was about the best way to remove linear terms in the analysis, and the second was concerned with normalization of data through division by the intra-height standard deviations.

The method used for linear trend removal was simply to compute the least square linear regression coefficients of x on h and y on h separately and then to remove the linear trend lines from x and y to form zero-mean residuals u and v. The question was then raised as to whether a better method was available such as equidistant end points. This is a difficult question to answer on a mathematical basis. A surface could be removed as readily as trend lines. The answer seems only to be found by trial and error. If all harmonics were used for prediction, the effect of different methods would be minimized, and elaborate and time-consuming comparisons would not seem to be justified at this time.

The 4-term CPVE for the normalized monthly Cape Kennedy data was consistently larger than for the non-normalized daily Montgomery soundings in MWP-I, either because of smoothing by monthly averaging or because of normalization. No intra-level standard deviation pertaining soley to an individual sounding is available. Reciprocal intra-level standard deviations may be used as a weighting function, but which of the available climatological values is best for filtering or smoothing is not known. This question should be answered as a refinement to the first or subsequent predicting methods by experimenting with various intra-level standards deviations based on hours, days, weeks, months, or a year. These would also be a function of the time scale of the prediction desired.

The degree to which averaging and normalizing of wind data stabilizes sequential Fourier representations is of interest. To examine this question, comparisons were made using the complex Fourier coefficients from the 24-point, normalized Cape Kennedy monthly averages given in MWP-I. As a crude estimate of the similarity between profile representations, a number of Spearman rank correlations along with a W (coefficient of concordance) value were computed. Considering the degree of agreement among all 12 months, the value W = 0.685 (r_{sav} = 0.656) was found, which is significant at less than the 0.001 level. This shows that a surprising amount of similarity exists between all 12 monthly Fourier representations.

As the monthly data were averaged over a number of years, the twelve months can be considered to be circular for lag comparisons. The Spearman values were computed for twelve comparisons of one, two and three month lags. Examples would be for lag one Dec.-Jan., Jan.-Feb., ..., Nov.-Dec.; lag two Dec.-Feb., ..., Nov.-Jan.; lag three Dec.-Mar., ..., Nov.-Feb.

The frequency of occurrence of r_s values for the monthly data is given in Table 11. All 36 coefficients were significant at the 0.01 level.

TABLE 11

Frequency of Rank Correlations for Monthly Cape Kennedy Data

r_s	LAG I	II	III
$\geq$ 0.80	2	0	1
0.70-0.79	4	6	2
0.60-0.69	4	4	7
0.50-0.59	2	2	2

The W and r_s values for Cape Kennedy are reasonably large and indicate a relationship between the 12 monthly Fourier representations. These W and r_s values are larger than those found for the Montgomery data, indicating that monthly representations are related and in a better fashion than the long Montgomery soundings in intervals or 6-hour sequences. This is probably due in part to the smoothing by taking averages, but some is probably due to normalization.

6. CONCLUSIONS (II)

The method of augmented Fourier polynomials has been used to describe the wind mathematically for different seasons, basic wind repre-

sentations (velocity and momentum), soundings of 34 kilometers, 24-point overlapping intervals within a long sounding, and sequences of 6-hour soundings. The results presented show the versatility of the method in a variety of situations.

Although the conclusions are based on a limited number of samples and are at best only indications, most of the results of analysis are reasonably consistent over all samples and are valuable as guides for prediction. This summary will be mostly in terms of a linear plus four harmonic (reduced harmonic) representation, unless otherwise noted.

The CPVE (cumulative percent variance explained) and the four most important harmonics of the new soundings analyzed were quite similar to the findings in MWP-I. The CPVE, although a little lower generally, was consistently in the 82-95% range. The important harmonic numbers were again found to be the 1st, 23rd, 2nd, and 22nd. However, the new data indicated that for summer the 22nd harmonic should be replaced by the 21st. The ability of four important harmonics given above plus a linear term to explain a large percentage of the variance was shown in the section on arbitrary selection, where the CPVE by the selected harmonics was only about 1 percent variance explained less than that given by the actual four most important harmonics plus the linear term.

The momentum representation of the wind showed consistently larger CPVE than the velocity in both winter and summer. Seasonal comparisons of CPVE are a little confused. In the two long soundings, winter had a larger CPVE, while in the 6-hour sequences, the summer values were larger than the winter for the new data, but both were smaller than the CPVE values for the winter sequence given in MWP-I. Probably, based on these results and the configuration of the winter wind profiles in general, winter will be better for prediction using a reduced harmonic representation.

The effect of profile length on CPVE was examined. Indications from the data are that CPVE vary inversely with the length of the interval taken. This also follows logically when the frequency content of the original sounding is considered.

In the decomposition of a long sounding into 24-point overlapping intervals, the CPVE was found to vary within the 82-95% range. The individual importance of the harmonics also varied from interval to interval, but in general the complex Fourier coefficients showed a good agreement when considered over all intervals, showing that the interval chosen is not usually critical. However, indications were given that the representations found for the intervals including the surface and first kilometer did not agree well with the other interval representations, suggesting that the surface and first kilometer be excluded from soundings for prediction.

The 6-hour sequences showed considerable variation in CPVE and

representation from sounding to sounding. Although the first four important harmonics were usually the same, they varied in the relative magnitude of importance within each profile. The complete sets of Fourier coefficients did not agree very well from sounding to sounding in a sequence, suggesting some problems in prediction may be encountered.

The effect of averaging and normalizing on a sounding also was examined. Comparisons with the monthly Cape Kennedy data showed that averaging and normalizing does have a stabilizing effect on the complex Fourier representation.

All this preliminary work should prove valuable for the development of a set of regression equations to predict wind profiles. The method of regression and ultimately prediction will follow.

APPENDIX II-A

Long Soundings from Montgomery, Alabama

	1000Z 8 Jun 56			0300Z 12 Feb 56		
km	Density	Xcomp	Ycomp	Density	Xcomp	Ycomp
0	1.202	0.000	0.000	1.254	1.532	1.286
1	1.071	-0.436	-4.981	1.140	7.518	-2.736
2	0.984	0.696	-4.951	1.011	12.840	-2.034
3	0.879	2.509	-6.535	0.910	17.669	3.435
4	0.797	3.527	-4.854	0.817	22.092	9.378
5	0.722	1.658	-1.118	0.737	29.670	11.988
6	0.652	1.042	-5.909	0.665	27.851	19.502
7	0.587	2.924	-9.563	0.592	31.947	22.370
8	0.528	-3.708	-11.413	0.530	38.105	22.000
9	0.475	-7.452	-15.279	0.464	40.189	31.399
10	0.425	-7.458	-21.747	0.405	40.873	36.802
11	0.378	-7.321	-16.444	0.347	45.962	38.567
12	0.333	-4.918	-18.353	0.298	37.900	34.126
13	0.291	-3.598	-22.717	0.260	38.184	38.184
14	0.246	-5.198	-24.454	0.225	45.962	38.567
15	0.211	0.000	-23.000	0.194	25.916	27.791
16	0.180	-1.099	-20.971	0.168	34.648	34.648
17	0.153	-1.706	-13.986	0.143	23.783	28.343
18	0.130	0.471	-8.988	0.121	9.506	14.094
19	0.110	-1.042	-5.909	0.103	8.756	14.572
20	0.093	-6.252	-6.474	0.088	1.710	4.698
21	0.078	-5.563	-2.248	0.074	1.307	14.943
22	0.066	-2.971	0.418	0.063	2.877	-2.779
23	0.056	-3.990	-0.279	0.054	-3.473	-3.597
24	0.048	-6.894	-5.785	0.046	-0.070	1.999
25	0.041	-10.064	-6.536	0.039	4.881	10.963
26	0.034	-11.370	-6.303	0.033	1.035	3.864
27	0.029	-11.820	-9.235	0.028	0.345	1.813
28	0.025	-11.326	-8.229	0.024	3.064	-2.571
29	0.021	-9.830	-6.883	0.020	2.158	-2.084
30	0.018	-7.947	-4.225	0.017	15.035	5.472
31	0.016	-9.456	-7.388	0.015	22.825	7.416
32	0.013	-10.466	-13.396	0.013	12.364	4.017
33	0.012	-11.437	-17.612	0.011	12.364	4.017

MATHEMATICAL WIND PROFILES
PART III

SUMMARY

Serially complete 6-hour wind observations from the surface to 27 km over Cape Kennedy during 1962 were used to compute 5200 serial correlations of wind integrated over 7 km layers. The correlations are for four representations of the wind, four atmospheric zones and their sum, thirteen calendar intervals, four observation times separately and all four combined, and four time lags. The correlations have been used to formulate criteria for wind profile prediction by the augmented Fourier polynomials, which represent mathematically a vertical two-dimensional wind profile.

1. INTRODUCTION (III)

Augmented Fourier polynomials for representing mathematically a vertical wind profile, and various properties of this method, are described in "Mathematical Wind Profiles, Part I and II". Before such polynomials could be used for a wind prediction model, answers were required to two questions:

a. Which basic physical representation, speed or momentum, and in scalar or vector form, is best for prediction?

b. How many past soundings should be used to predict one future sounding?

To answer these questions, serial correlations were computed for scalar speed, vector speed, scalar momentum, and vector momentum over time lags of 6, 12, 18, and 24 hours. Rather than for individual levels, integrated air movement over layers (zones) 7 km thick was used.

As basic wind data, a serially complete set of winds from Cape Kennedy for the year 1962 were supplied on magnetic tape by Mr. O. E. Smith of NASA, George C. Marshall Space Flight Center, Huntsville, Alabama. It provided velocity V and direction θ at 1 km intervals from 0 through 27 km at 6-hour intervals (0000, 0600, 1200, and 1800 GMT) on all 365 days. For correlation computation, each wind sounding is denoted as T_i where i = 1, 2,..., 1460, in sequence from 0000 GMT on 1 Jan. 1962 (i = 1) through 1800 GMT on 31 Dec. 1962 (i = 1460).

To obtain the vector representation, values for each level (h) on each sounding (i) were transformed from polar (V_h, θ_h) to cartesian (x_h, y_h) coordinates by

$$x_h = - V_h \sin \theta_h, \qquad y_h = - V_h \cos \theta_h .$$

For each of 5 atmospheric zones (four of 7 levels each, one for all levels), four different physical representations of the wind were used:

Scalar Speed: $G = \Sigma_h V_h$

Vector Speed: $G' = \left[(\Sigma_h x_h)^2 + (\Sigma_h y_h)^2 \right]^{1/2}$

Scalar Momentum: $D = \Sigma_h q_h V_h$

Vector Momentum: $D' = \left[(\Sigma_h q_h x_h)^2 + (\Sigma_h q_h y_h)^2 \right]^{1/2}$

Here q_h is the atmospheric density at height h on the ith sounding.

The index of summation for h can be changed appropriately to define the different atmospheric zones. Three of the zones, 7 to 13, 14 to 20, and 21 to 27 km, may be considered as 7 km thick, e.g. 6.5 to 13.5 km. The lowermost zone, 0 to 6 km, is at most 6.5 km thick, and the fifth zone, 0 to 27 km, is 27.5 km thick.

Density values for individual soundings were not included in the basic data tape and were not otherwise readily available to compute momentum. Instead, average monthly density profiles (0-27 km) from Cape Kennedy[5] were examined. The difference between the maximum and minimum density at each height level over all months ranged from 0.6 to 5.9 per cent of maximum density; the average difference for all heights was about 3.2 per cent, much smaller than the difference in densities at the surface and 27 km, which differ by a factor of 40. Because any average monthly density profile would yield approximately the same result, momentum for each sounding was computed from the December average density profile, a typical high density profile (Table 1).

TABLE 1

AVERAGE MONTHLY DENSITY (g/cm^3) FOR DECEMBER AT CAPE KENNEDY

From "Atlantic Missile Range Reference Atmosphere for Cape Kennedy, Florida (Part 1)", IRIG Document 104-63.

Zone 1		Zone 2		Zone 3		Zone 4	
h	q_h	h	q_h	h	q_h	h	q_h
0	1.2097	7	0.5885	14	0.2504	21	0.0777
1	1.0982	8	0.5285	15	0.2150	22	0.0652
2	0.9895	9	0.4736	16	0.1850	23	0.0554
3	0.8915	10	0.4230	17	0.1576	24	0.0470
4	0.8044	11	0.3766	18	0.1328	25	0.0398
5	0.7251	12	0.3328	19	0.1110	26	0.0341
6	0.6534	13	0.2902	20	0.0926	27	0.0293

---------- Zone 5 ----------

After scalar and vector speeds and momenta G, G', D, and D' were computed over each of the five atmospheric zones of each sounding, they were divided into 13 equal time intervals, representing calendar intervals, for serial correlation. Each interval contains 28 days of 112 soundings each (Table 2). In each, one day (four soundings) of overlap from the previous interval was included for continuity of serial correlations; for the initial interval, the 1st of January was the overlap day. Thus the initial interval is composed of 112 soundings from 2 Jan. through 29 Jan., with the four soundings from 1 Jan. for overlap, and the last interval is composed of 112 soundings from 4 Dec. through 31 Dec., with 3 Dec. for overlap.

TABLE 2

CALENDAR INTERVALS USED FOR SERIAL CORRELATIONS

Intervals go from the 0000 GMT observation on the begin date to the 1800 GMT observation on the end date; the first date includes one day of overlap.

Interval	Begin	End	Record Numbers
0	Jan 1	Jan 29	1 - 116
1	Jan 29	Feb 26	113 - 228
2	Feb 26	Mar 26	225 - 340
3	Mar 26	Apr 23	337 - 452
4	Apr 23	May 21	449 - 564
5	May 21	Jun 18	561 - 676
6	Jun 18	Jul 16	676 - 788
7	Jul 16	Aug 13	785 - 900
8	Aug 13	Sep 10	897 - 1012
9	Sep 10	Oct 8	1009 - 1124
10	Oct 8	Nov 5	1121 - 1236
11	Nov 5	Dec 3	1233 - 1348
12	Dec 3	Dec 31	1345 - 1460

Time lags of 6, 12, 18, and 24 hours were denoted $t = 1, 2, 3, 4$ for computation of serial correlations. One further stratification for each representation, atmospheric zone, and calendar interval was made on the basis of observing times: 0000, 0600, 1200, and 1800 separately, and all times together. This provides the serial correlation of the 0000 observation with that of 0600, 1200, 1800, and 0000 (next day) for time lags of 6, 12, 18, and 24 hours. Similar relationships are found in terms of the 0600, 1200, and 1800 observations, while all times gives the simple lag correlations irrespective of observing time.

2. CORRELATION COMPUTATIONS

All serial correlation computations used a slightly modified formula:

$$R_t = \mathrm{Cov}\,(x_i, x_{i+t}) \left[\mathrm{Var}\,(x_i)\ \mathrm{Var}\,(x_{i+t})\right]^{-1/2} \qquad (2.1)$$

$$= \frac{(N - A)\,(\Sigma\, x_i\, x_{i+t}) - (\Sigma\, x_i)\,(\Sigma\, x_{i+t})}{\left[(N - A)\,\Sigma\, x_i^2 - (\Sigma\, x_i)^2\right]^{1/2} \left[(N - A)\,\Sigma\, x_{i+t}^2 - (\Sigma\, x_{i+t})^2\right]^{1/2}}$$

where $A = 1$ when $N = 29$ and $A = 4$ when $N = 116$.

The one-day overlap provides four additional values for possible multiplication. Therefore, the number of multiplications $(N - A)$ for fixed N is the same for all four time lags. The summations in (2.1) have been appropriately indexed so that the same multiplicative pair is not included in two consecutive calendar intervals. This allows for simple recombining of calendar intervals to obtain serial correlations over longer time intervals in multiples of the original interval length. To accomplish the recombination if needed, the output includes the necessary sums, sums of squares, and cross products.

Each sounding was indexed as T_i, where $i = 1, 2, 3, \ldots, 1460$. Thus, T_1, T_2, T_3, and T_4, (treated as overlap for this calendar interval) represent the 0000, 0600, 1200, 1800 GMT soundings on 1 Jan. 1962; T_5, T_6, T_7, and T_8, represent the same sequence on 2 Jan., and T_{1457}, T_{1458}, T_{1459}, and T_{1460}, the four observations on 31 Dec. 1962. The same indexing was used on all four wind representations G_i, G_i', D_i, D_i'.

A particular wind representation in a particular atmospheric zone may be denoted as x. A dummy index K on x provides the appropriate multiplicative pairs for correlations associated with each observing time. The covariance is given by

$$\mathrm{Cov}\,(x_i, x_{i+L}) = \frac{1}{28} \sum_{i=1+E}^{29-F} x_{K+(i-1)4}\ x_{K+(i-1)4+L} \qquad (2.2)$$

$$- \frac{1}{28} \sum_{i=1+E}^{29-F} x_{K+(i-1)4}\ \frac{1}{28} \sum_{i=1+E}^{29-F} x_{K+(i-1)4+L}$$

$K = 1, 2, 3, 4$ gives correlations for the 0000, 0600, 1200, 1800 GMT observations, $L = 1, 2, 3, 4$ gives correlations for each K for time lags of 6, 12, 18, and 24 hours;

$$E = 1\,, \quad F = 0 \qquad \text{for} \qquad K + L \leq 4$$

$$E = 0\,, \quad F = 1 \qquad \text{for} \qquad K + L \geq 5\,.$$

For the correlation based on the 0000 observing time, the multiplicative pairs for the cross-product term, (VARIABLE Z in the printout, Fig. 1)

$$\sum_{i=1+E}^{29-F} x_{1+(i-1)4}\; x_{1+(i-1)4+L} \tag{2.3}$$

for the lags $L = 1, 2, 3, 4$ are:

Lag 1: $\Sigma\; x_5 \cdot x_6 + x_9 \cdot x_{10} + \ldots + x_{113} \cdot x_{114}$

Lag 2: $\Sigma\; x_5 \cdot x_7 + x_9 \cdot x_{11} + \ldots + x_{113} \cdot x_{115}$

Lag 3: $\Sigma\; x_5 \cdot x_8 + x_9 \cdot x_{12} + \ldots + x_{113} \cdot x_{116}$

Lag 4: $\Sigma\; x_1 \cdot x_5 + x_5 \cdot x_9 + \ldots + x_{109} \cdot x_{113}$.

The first and last multiplicative pairs for the first calendar interval (M = 0) for the four observing times are given in Table 3. From this table the first and last pairs for other calendar intervals can be obtained by adding 112(M) to the subscripts of each term (M = 0, 1, 2, ..., 12). The sequence of multiplicative pairs for fixed M is then found by sccessively adding 4 to each term in the pair, starting with the first pair.

TABLE 3

FIRST AND LAST MULTIPLICATIVE PAIR IN THE CROSS-PRODUCT TERM FOR THE FIRST CALENDAR INTERVAL ($M = 0$; $x_1, x_2, \ldots, x_{116}$)

Lag	First Pair			
	0000	0600	1200	1800
1	$x_5 \cdot x_6$	$x_6 \cdot x_7$	$x_7 \cdot x_8$	$x_4 \cdot x_5$
2	$x_5 \cdot x_7$	$x_6 \cdot x_8$	$x_3 \cdot x_5$	$x_4 \cdot x_6$
3	$x_5 \cdot x_8$	$x_2 \cdot x_5$	$x_3 \cdot x_6$	$x_4 \cdot x_7$
4	$x_1 \cdot x_5$	$x_2 \cdot x_6$	$x_3 \cdot x_7$	$x_4 \cdot x_8$

Lag	Last Pair			
	0000	0600	1200	1800
1	$x_{113} \cdot x_{114}$	$x_{114} \cdot x_{115}$	$x_{115} \cdot x_{116}$	$x_{112} \cdot x_{113}$
2	$x_{113} \cdot x_{115}$	$x_{114} \cdot x_{116}$	$x_{111} \cdot x_{113}$	$x_{112} \cdot x_{114}$
3	$x_{113} \cdot x_{116}$	$x_{110} \cdot x_{113}$	$x_{111} \cdot x_{114}$	$x_{112} \cdot x_{115}$
4	$x_{109} \cdot x_{113}$	$x_{110} \cdot x_{114}$	$x_{111} \cdot x_{115}$	$x_{112} \cdot x_{116}$

When all observing times are used the appropriate pairs in the cross-product summation for $M = 0$ are

Lag 1: $\Sigma\ x_4 \cdot x_5 + x_5 \cdot x_6 + \ldots + x_{115} \cdot x_{116}$

Lag 2: $\Sigma\ x_3 \cdot x_5 + x_4 \cdot x_6 + \ldots + x_{114} \cdot x_{116}$

Lag 3: $\Sigma\ x_2 \cdot x_5 + x_3 \cdot x_6 + \ldots + x_{113} \cdot x_{116}$

Lag 4: $\Sigma\ x_1 \cdot x_5 + x_2 \cdot x_6 + \ldots + x_{112} \cdot x_{116}$

The first and last pairs for other intervals are found by adding 112(M) to each term of the pair.

3. CORRELATION SIGNIFICANCE

The significance of serial correlations is much more difficult to assess than that of ordinary product-moment correlations. Except when the true correlation $\rho = 0$, the usual (linear) product moment r_{xy} gives a biased estimate of ρ, and the sampling distribution of r_{xy} about ρ is skewed and difficult to evaluate directly for significance tests.

The z' transformation, introduced by R. A. Fisher, is usually preferred for significance testing:

$$z' = \frac{1}{2} \ln \frac{1 + r}{1 - r} \tag{3.1}$$

The variable z' is asymptotically normally distributed, with mean and variance

$$\zeta' = \frac{1}{2} \ln \frac{1 + \rho}{1 - \rho} + \frac{\rho}{2 (n + 1)} \left[1 + \frac{1 + \rho^2}{8 (n - 1)} + \cdots \right]$$

$$\sigma^2_{z'} = \frac{1}{n - 1} \left[1 + \frac{4 - \rho^2}{2 (n - 1)} + \frac{176 - 21\rho^2 - 21\rho^4}{48 (n - 1)^2} + \cdots \right] \tag{3.2}$$

Very good approximations are obtained by using only the first term for the mean, and $1/(n - 3)$ for the variance, as long as only one sample correlation is being tested. But valid evaluation of a complete set of correlation coefficients for different lags requires not only the sampling distribution of R_t about ρ_t, but also the joint variation among $R_1, R_2, \ldots, R_p$.

Most of the results on significance testing of serial correlation have been in terms of large sample theory, for testing for serial dependence with R_1. General distributions, derived under appropriate assumptions, depend upon the unknown population values of ρ_t and, for covariance, also on all possible population partial correlation coefficients. For coefficients other than R_1, a test, based on a circular definition and other assumptions, uses an incomplete beta function. Most sampling distributions assume normality and stationarity, which may not be met in practice. Some problems can be avoided by assuming a theoretical form for ρ_t and testing the sample results for goodness of fit. If an autoregressive model is desired, modified regres-

sion methods can be used to test, provided not too much accuracy is desired.

The serial correlations reported here were computed to obtain a general evaluation of four representations of the wind, and of how many lags were to be used in the autoregressive prediction model. For more elaborate testing for comparative purposes, the standard z' transformation can probably be used for some significance tests and confidence intervals with reasonably accurate results, depending on the rigor needed. Hannan[6] suggests that R_1 can be tested by considering $(R_1 + 1/n)$ as an ordinary correlation coefficient from $(n + 3)$ observations.

Testing individual coefficients other than R_1 requires extreme care. A significance test of R_1 is merely a test of serial dependence among all observations, but a test of R_2 or higher order coefficients in an autoregressive model is equivalent to determining how much serial dependence exists between observations separated by more than one time step, and whether the dependence is large enough to be important in explaining (modeling) the process. Such a test is based on partial correlation coefficients, which represent the amount of dependence left for that particular lag after the effects (dependences) of all other lags on the process have been removed from the serial correlation coefficient. The lag correlation values computed for $t = 2$, 3, and 4 are not partial correlation coefficients, and should not be treated as such.

For direct evaluation of the lag 1 correlations, Table 4 presents approximate 95% confidence intervals for the true correlation coefficient based on correlations obtained from samples of 28 and 112 observations. These confidence intervals were computed from the z' transformation (3.1) by adding and subtracting two standard deviations: $z' \pm 0.14$, for 28 and 112 observations, respectively. Results are in excellent agreement with exact confidence intervals scaled from the charts in David's monograph[7], or interpolated from the tables therein.

Table 4 shows, for example, that in a sample of 28 observation pairs, the correlation must be greater than about 0.50 in order to be considered significantly different from 0. Almost all the lag 1 correlations well exceeded this value, and hence may be considered "significant". But, for 28 pairs, correlations of 0.80 and 0.90 cannot be considered as differing, since the confidence interval for each includes the other.

TABLE 4

95% CONFIDENCE INTERVALS FOR CORRELATION COEFFICIENTS BASED ON 28 AND 112 PAIRS

r	n = 28		n = 112	
	Lower	Upper	Lower	Upper
0.10	-0.29	0.46	-0.04	0.24
0.20	-0.20	0.54	0.06	0.33
0.30	-0.10	0.60	0.16	0.41
0.40	-0.02	0.68	0.27	0.51
0.50	0.15	0.74	0.39	0.60
0.55	0.22	0.77	0.45	0.64
0.60	0.28	0.80	0.50	0.68
0.65	0.36	0.83	0.56	0.73
0.70	0.44	0.85	0.62	0.77
0.75	0.52	0.88	0.69	0.80
0.80	0.60	0.91	0.74	0.85
0.82	0.64	0.92	0.77	0.86
0.84	0.68	0.92	0.79	0.88
0.86	0.71	0.93	0.82	0.89
0.88	0.75	0.94	0.85	0.91
0.90	0.79	0.95	0.87	0.92
0.91	0.81	0.96	0.88	0.93
0.92	0.83	0.96	0.90	0.94
0.93	0.85	0.97	0.91	0.95
0.94	0.87	0.97	0.92	0.95
0.95	0.89	0.98	0.94	0.96
0.96	0.91	0.98	0.95	0.97
0.97	0.94	0.99	0.96	0.98
0.98	0.96	0.99	0.97	0.98
0.99	0.98	1.00	0.99	0.99

M COUNT IS 0 RECORD NOS. RANGE FROM 0 TO 116 LISTED VARIABLE IS G 0-27 KM

VARIABLE Z

K/L	0000	0600	1200	1800	TOTAL
1	0.10343210E 08	0.10346401E 08	0.10401677E 08	0.10668038E 08	0.41759328E 08
2	0.10311355E 08	0.10307505E 08	0.10479768E 08	0.10578862E 08	0.41677490E 08
3	0.10286482E 08	0.10232036E 08	0.10371446E 08	0.10537264E 08	0.41427229E 08
4	0.10066864E 08	0.10103738E 08	0.10334603E 08	0.10523838E 08	0.41029043E 08

VARIABLE T

K/L	0000	0600	1200	1800	TOTAL
1 I	0.10400864E 08	0.10350984E 08	0.10426426E 08	0.11003385E 08	0.42181659E 08
I+1	0.10350984E 08	0.10426426E 08	0.10441025E 08	0.10400864E 08	0.41619300E 08
2 I	0.10400864E 08	0.10350984E 08	0.10745969E 08	0.11003385E 08	0.42501203E 08
I+1	0.10426426E 08	0.10441025E 08	0.10400864E 08	0.10350984E 08	0.41619300E 08
3 I	0.10400864E 08	0.10444909E 08	0.10745969E 08	0.11003385E 08	0.42595128E 08
I+1	0.10441025E 08	0.10400864E 08	0.10350984E 08	0.10426426E 08	0.41619300E 08
4 I	0.10310879E 08	0.10444909E 08	0.10745969E 08	0.11003385E 08	0.42505144E 08
I+1	0.10400864E 08	0.10350984E 08	0.10426426E 08	0.10441025E 08	0.41619300E 08

VARIABLE S

K/L	0000	0600	1200	1800	TOTAL
1 I	0.16509999E 05	0.16434000E 05	0.16568000E 05	0.16993999E 05	0.66505999E 05
I+1	0.16434000E 05	0.16568000E 05	0.16654000E 05	0.16509999E 05	0.66165999E 05
2 I	0.16509999E 05	0.16434000E 05	0.16758999E 05	0.16993999E 05	0.66696999E 05
I+1	0.16568000E 05	0.16654000E 05	0.16509999E 05	0.16434000E 05	0.66165999E 05
3 I	0.16509999E 05	0.16498999E 05	0.16758999E 05	0.16993999E 05	0.66762000E 05
I+1	0.16654000E 05	0.16509999E 05	0.16434000E 05	0.16568000E 05	0.66165999E 05
4 I	0.16435999E 05	0.16498999E 05	0.16758999E 05	0.16993999E 05	0.66688000E 05
I+1	0.16509999E 05	0.16434000E 05	0.16568000E 05	0.16654000E 05	0.66165999E 05

VARIABLE R

K/L	0000	0600	1200	1800	TOTAL
1	0.95283285E 00	0.93859994E 00	0.94759855E 00	0.95600712E 00	0.94656631E 00
2	0.84181976E 00	0.86691857E 00	0.86652746E 00	0.86708784E 00	0.85735642E 00
3	0.78137524E 00	0.72575954E 00	0.75342379E 00	0.73511537E 00	0.74636810E 00
4	0.56516032E 00	0.58817313E 00	0.62639092E 00	0.68484742E 00	0.61337812E 00

FIGURE 1

4. RESULTS

The 5200 serial correlations, and the sums of squares and cross-products from which they were derived, form a computer printout about six inches thick. One page of this output is reproduced in Figure 1, and the various entries thereon are identified in this Section.

The first row classifies the correlation. The "M COUNT IS" gives the calendar interval (see Table 2). The "RECORD NOS. RANGE FROM" gives the range of the soundings used for the correlation in terms of each sounding indexed from 1 through 1460. The "LISTED VARIABLE IS" gives the type of wind representation: G is scalar speed, G' vector speed, D scalar momentum, D' vector momentum. The last entry on the right, the atmospheric zone, is made by hand on the basis of the computational order; for a fixed M and representation, the atmospheric zones were computed in the order 0-6, 7-13, 14-20, 21-27, and 0-27.

Column headings refer to the time of observation (0000, 0600, 1200, 1800 GMT). The "TOTAL" heading is used for all four observing times considered together.

Row headings are the lag units, where L = 1, 2, 3, 4 refer to time lags of 6, 12, 18, and 24 hours, respectively. The I entry refers to the quantity being summed, with the first value used in the correlation and excluding the last, while the I + 1 entry refers to the first being excluded and the last included.

The first three matrices (Fig. 1) give the values used to compute the serial correlation coefficients in the fourth matrix:

VARIABLE Z gives the cross-product term $\Sigma\, x_i\, x_{i+t}$.

VARIABLE T gives the sums of squares $\Sigma\, x_i^2$ and $\Sigma\, x_{i+t}^2$.

VARIABLE S gives the sums $\Sigma\, x_i$ and $\Sigma\, x_{i+t}$.

VARIABLE R gives the serial correlation coefficients R_t of (2.1).

For the column headed TOTAL, (N-A) is 112 for each lag. For the other four columns, (N-A) is 28 for each lag.

In almost all cases, the correlation coefficients for 6 hour lags were considerably higher than for 12, 18, or 24 hour lags. For example, correlations for the various lags for vector speed for the entire 27 km profile, considering all observation times, were:

Calendar Interval:	0	1	2	3	4	5	6	7	8	9	10	11	12
6-hr lag:	.95	.92	.93	.96	.93	.85	.94	.91	.96	.93	.97	.90	.96
12-hr lag:	.87	.81	.88	.89	.81	.71	.86	.79	.90	.84	.92	.74	.88
18-hr lag:	.76	.72	.84	.77	.67	.55	.76	.70	.82	.73	.85	.55	.80
24-hr lag:	.63	.63	.81	.64	.54	.45	.66	.61	.73	.64	.79	.37	.72

Correlations for lags 2, 3, and 4, corresponding to 12, 18, and 24 hours, were not evaluated rigorously. They were generally smaller than the 6-hour lag correlations, and were discarded without further testing to determine whether the differences were significant because of the evaluation difficulties discussed in the previous Section. Further examination was restricted to the 6-hour lag values.

5. COMBINED INTERVALS

Correlations for 6-hour lags for combinations of two and more calendar intervals were computed to demonstrate the feasibility and results of such computation. The one-day overlaps provided at the start of each interval insured that the same multiplicative pair was used in only one calendar interval.

Values for the cross-products, sums of squares, and sums (matrices Z, T, and S in Fig. 1) for one calendar interval can be added to the corresponding values in the next interval to provide a serial correlation coefficient for the combined period. Such combination can be extended to as many intervals as desired, with (N-A) in Eq. (2.1) taken as the number of combined intervals multiplied by 28 or 112; the variances and covariance are obtained from the sums of the corresponding values for each interval.

Combined coefficients were computed only for the scalar speed representation for the entire 0-27 km zone, for all observations ("total"). Computations were for the first two, first three, etc., intervals up to the entire year of 13 intervals. Coefficients for 6-hour lag of scalar wind for the entire profile, for each interval separately and for the combined interval are:

Interval:	0	1	2	3	4	5	6	7	8	9	10	11	12
Separate:	.95	.92	.91	.94	.88	.90	.81	.86	.87	.92	.97	.92	.95
Combined:	.95	.94	.93	.94	.96	.97	.91	.98	.98	.98	.98	.98	.98
Speed:	21	19	23	17	11	9	10	9	9	9	12	15	20

The last line gives the mean speed, in knots, for the interval, found by averaging the wind speeds in each profile over all heights.

Correlations for combined intervals (2nd line) are greater than the average of the correlations of the individual intervals (1st line) because of the differences in the means for the intervals (3rd line). This effect can be examined most readily by computing the correlation for a combination of two intervals, A and B, each of length $N = 112$. For lag 1 correlation, the means and variances for the first N observations do not differ materially from those for the N observations from 2 to $N + 1$, and the covariance between the first N and second N observations is substantially the same as that between the observations from 2 to $N + 1$ and from $N + 2$ to $2N + 1$.

Assuming these equalities, straightforward but tedious algebra (Appendix A) gives the lag 1 correlation over the two intervals in terms of the average, $\bar{r}$, of the correlations r_A and r_B over the two intervals separately, and the corresponding variances S_A^2 and S_B^2, and the difference Δ between their means:

$$R_{1,AB} = \bar{r} + \frac{1 - \bar{r}}{1 + 2\,(S_A^2 + S_B^2)\,\Delta^{-2}} + \frac{(r_A - r_B)(S_A^2 - S_B^2)}{\Delta^2 + 2\,(S_A^2 + S_B^2)} \qquad (5.1)$$

As long as $\bar{r} > 0$, the second term is positive. The third term vanishes if either the correlations are equal or the variances are equal, and

$$R_{1,AB} = \bar{r} + \frac{(1 - \bar{r})}{1 + 4S^2/\Delta^2} \quad . \qquad (5.2)$$

In this special case, the correlation for the combined sample is never less than the average, $\bar{r}$, of the correlations over A and B separately, and is greatest when $(S/\Delta)^2$ is smallest, i.e. when $\Delta^2 \gg S^2$. Thus when the difference between the sample means is much larger than their common standard deviation, the correlation over the two samples will exceed the

the average correlation. Only when the difference between the means, Δ, is zero will the combined correlation be no greater than the average. Even under more general conditions, without equality of variances, these tendencies still are present, as can be shown by evaluation of the second and third terms of (5.1).

The square of the (linear) correlation coefficient represents the relative reduction in variance of one variable when it is predicted by (linear) regression on the other. The increase in correlation achieved by combining samples with different means thus represents an increase in prediction accuracy, but only with respect to the unconditional variance, which is the combined variance over the two samples. This is greater than the mean variance over the two samples by one-fourth the squared difference of the two sample means, and hence the apparent increase in prediction accuracy is illusory. Samples should not be combined for correlation, or for regression prediction, if their means differ significantly.

The usual t - test for equality of means in two normal samples, each of size N, involves

$$N^{1/2} \frac{\bar{x}_A - \bar{x}_B}{(S_A^2 + S_B^2)^{1/2}} = t_f \quad , \tag{5.3}$$

where the number of degrees of freedom is given by

$$f = (N+1) \frac{(S_A^2 + S_B^2)^2}{S_A^4 + S_B^4} - 2 \quad . \tag{5.4}$$

Hence $t^2/N = \Delta^2/(S_A^2 + S_B^2)$ and (5.1) becomes

$$R_{1,AB} = \bar{r} + \frac{1 - \bar{r}}{1 + 2Nt^{-2}} + \frac{S_A^2 - S_B^2}{S_A^2 + S_B^2} \quad \frac{r_A - r_B}{2 + t^2/N} \quad . \tag{5.5}$$

For large samples, the 0.05% value of t is approximately 2. Even if the difference in the sample means is not quite large enough to warrant rejection of a hypothesis of equality, combining the samples will increase the correlation, and hence the apparent accuracy of prediction.

For purposes of wind profile prediction, combining samples seems unwarranted. The period of interest should be broken into intervals just large enough for valid means, variances, and covariances to be computed, unless no trend whatever is expected in the means.

6. CONCLUSIONS (III)

Tables 5 to 9 present all 6-hour lag serial correlation coefficients for the five atmospheric zones. In general, the correlations were quite high, and all were significantly greater than zero at the 95% level. All correlations presented in the tables ranged from 0.58 to 0.98.

As a quick method of assessing which representation was best, the correlations for the four representations were ranked on the basis of magnitude for each atmospheric zone and time of observation over each of the 13 calendar intervals. Vector speed consistently ranked first or second more times for all observing times and atmospheric zones, except in the 14-20 km zone. In this zone, vector momentum, the next highest in ranking, was higher in the first and second ranking for all observing times.

The work presented in this report was directed toward finding the best wind representation and number time lags to be used for prediction. However, many other possibilities exist for using the data. Some of these uses have been touched on in an exploratory manner. The most important are various comparisons with atmospheric zones and observing times, the decay rate of the lag correlations and the behavior of the serial correlations with variable interval length.

Ultimately more work must be performed on the best calendar interval for prediction. Some of the summer and winter calendar intervals might be combined into longer intervals. Correlations for many different combinations of interval length (in units of 28 days) can be computed from the basic data.

In most cases, the correlation coefficients for vector speed are larger than those for vector momentum, scalar speed, or scalar momentum. The 6-hour lag correlations are consistently larger than those for the 12, 18, and 24 hour lags. Based on this evidence, the first prediction model will be in terms of vector speed and the previously observed sounding.

APPENDIX III-A

SERIAL CORRELATIONS FOR COMBINED INTERVALS

Serial correlation coefficients computed for combinations of two or more time intervals tend to be greater than the average of the serial correlations computed for each interval separately. This increase depends primarily on the differences between the means for the various intervals. It can be examined most readily by computing the correlation for a combination of two intervals, A and B , (With N = 112, addition of 3 to all summation limits will make them correspond to those implied by Eqs. (2.1) and (2.2), e.g. "Lag 1" in Table 3.)

Because $R_{1,AB} = \text{cov}_{AB}(x_i\, x_{i+1}) \left[\text{var}(x_i)\ \text{var}(x_{i+1})\right]^{1/2}$, the numerator and denominator are developed separately before combination.

$$\text{cov}_{AB}(x_i\, x_{i+1}) = \frac{1}{2N}\sum_{i=1}^{2N}(x_i\, x_{i+1}) - \frac{1}{2N}\sum_{i=1}^{2N}x_i \quad \frac{1}{2N}\sum_{i=1}^{2N}x_{i+1}$$

$$= \frac{1}{2}\left[\frac{1}{N}\sum_{1}^{N}x_i\, x_{i+1} + \frac{1}{N}\sum_{N+1}^{2N}x_i\, x_{i+1}\right] \qquad \text{(A.1)}$$

$$- \frac{1}{4}\left[\frac{1}{N}\sum_{1}^{N}x_i + \frac{1}{N}\sum_{N+1}^{2N}x_i\right]\left[\frac{1}{N}\sum_{1}^{N}x_{i+1} + \frac{1}{N}\sum_{N+1}^{2N}x_{i+1}\right]$$

Adding and subtracting the means,

$$\bar{x}_A = \frac{1}{N}\sum_{1}^{N}x_i\ , \quad \bar{x}_{A+} = \frac{1}{N}\sum_{2}^{N+1}x_i = \frac{1}{N}\sum_{'1}^{N}x_{i+1} \qquad \text{(A.2)}$$

with $\bar{x}_B$ and $\bar{x}_{B+}$ defined similarly, provides an expression involving variances and covariances:

$$\text{cov}_{AB}(x_i\ x_{i+1}) = \frac{1}{2}\left(\frac{1}{N}\sum_{1}^{N} x_i\ x_{i+1} - \bar{x}_A\,\bar{x}_{A+} + \frac{1}{N}\sum_{N+1}^{2N} x_i\ x_{i+1} - \bar{x}_B\,\bar{x}_{B+}\right)$$

$$+ \frac{1}{2}\left(\bar{x}_A\,\bar{x}_{A+} + \bar{x}_B\,\bar{x}_{B+}\right) - \frac{1}{4}\left(\bar{x}_A\,\bar{x}_{A+} + \bar{x}_A\,\bar{x}_{B+} + \bar{x}_{A+}\bar{x}_B + x_{A+}x_{B+}\right)$$

$$= \frac{1}{2}\left(\text{cov}_A + \text{cov}_B\right) + \frac{1}{4}\left(\bar{x}_A\,\bar{x}_{A+} - \bar{x}_A\,\bar{x}_{B+} - \bar{x}_{A+}\bar{x}_B + \bar{x}_B\,\bar{x}_{B+}\right)$$

$$= \frac{1}{2}\left(\text{cov}_A + \text{cov}_B\right) + \frac{1}{4}\left(\bar{x}_A - \bar{x}_B\right)\left(\bar{x}_{A+} - \bar{x}_{B+}\right).$$

$$= \frac{1}{2}\left(\text{cov}_A + \text{cov}_B\right) + \frac{1}{4}\ \Delta_{AB}\,\Delta_{A+B+} \qquad \text{(A.3)}$$

where $\Delta_{AB} = \bar{x}_A - \bar{x}_B$ and $\Delta_{A+B+} = \bar{x}_{A+} - x_{B+}$. The difference between these differences is

$$\Delta_{AB} - \Delta_{A+B+} = \frac{1}{N}\left(x_1 + \sum_{2}^{N} x_i - x_{N+1} - \sum_{N+2}^{2N} x_i\right) - \frac{1}{N}\left(\sum_{2}^{N} x_i + x_{N+1} - \sum_{N+2}^{2N} x_i - x_{2N+1}\right)$$

$$= \frac{1}{N}\left(x_1 - 2x_{N+1} + x_{2N+1}\right) \qquad \text{(A.4)}$$

On the average, this term is quite small, so the two differences may be equated with negligible error, and

$$\text{cov}_{AB}(x_i\ x_{i+1}) \doteq \frac{1}{2}\left(\text{cov}_A + \text{cov}_B\right) + \frac{1}{4}\ \Delta_{AB}^2 \qquad \text{(A.5)}$$

Similar development for the variances gives

$$S_{AB}^2 = \frac{1}{2N}\sum_{1}^{2N} x_i^2 - \left(\frac{1}{2N}\sum_{1}^{2N} x_i\right)^2$$

$$= \frac{1}{2}\left[\frac{1}{N}\sum_{1}^{N} x_i^2 - (\bar{x}_A)^2 + \frac{1}{N}\sum_{N+1}^{2N} x_i^2 - (\bar{x}_B)^2\right]$$

$$+ \frac{1}{4}\left[2(\bar{x}_A)^2 + 2(\bar{x}_B)^2\right] - \frac{1}{4}\left[(\bar{x}_A)^2 + 2\bar{x}_A\bar{x}_B + (\bar{x}_B)^2\right]$$

$$= \frac{1}{2}\left(S_A^2 + S_B^2\right) + \frac{1}{4}\left(\bar{x}_A - \bar{x}_B\right)^2$$

$$= \frac{1}{2}\left(S_A^2 + S_B^2\right) + \frac{1}{4}\Delta_{AB}^2 \qquad \text{(A.6)}$$

Likewise, $S_{A+B+}^2 = \frac{1}{2}\left(S_{A+}^2 + S_{B+}^2\right) + \frac{1}{4}\Delta_{A+B+}^2$. Because

$$\bar{x}_A + \bar{x}_{A+} = 2\bar{x}_A - (x_1 - x_{N+1})/N$$

$$\bar{x}_A - \bar{x}_{A+} = (x_1 - x_{N+1})/N$$

$$(\bar{x}_A)^2 - (\bar{x}_{A+})^2 = (\bar{x}_A + \bar{x}_{A+})(\bar{x}_A - \bar{x}_{A+})$$

$$= \left[2\bar{x}_A - (x_1 - x_{N+1})/N\right]\left[x_1 - x_{N+1}\right]/N \qquad \text{(A.7)}$$

the difference in variances is

$$S_A^2 - S_B^2 = \left[\frac{1}{N}\sum_1^N x_i^2 - (\bar{x}_A)^2\right] - \left[\frac{1}{N}\sum_1^N x_{i+1}^2 - (\bar{x}_{A+})^2\right]$$

$$= \frac{1}{N}\left(x_1^2 + \sum_2^N x_i^2 - \sum_2^N x_i^2 - x_{N+1}^2\right) - \left[(\bar{x}_A)^2 - (\bar{x}_{A+})^2\right]$$

$$= \frac{1}{N}(x_1 - x_{N+1})(x_1 + x_{N+1}) - \left[2\bar{x}_A - (x_1 + x_{N+1})/N\right](x_1 - x_{N+1})/N$$

$$= \frac{1}{N}(x_1 - x_{N+1})\left[(x_1 + x_{N+1}) - 2\bar{x}_A + (x_1 - x_{N+1})/N\right]$$

$$= N^{-2}(x_1 - x_{N+1})\left[(N+1)x_1 - 2N\bar{x}_A + (N-1)x_{N+1}\right]$$

$$= N^{-2}(x_1 - x_{N+1})\left[(N+1)(x_1 + x_{N+1}) - 2(x_1 + \ldots + x_{N+1})\right] \quad (A.8)$$

On the average, the first difference is small and the second one still smaller. Division by N^2 makes the whole expression negligible, and the difference in variances may be neglected.

Consequently, $S_A^2 \doteq S_{A+}^2$ and $S_B^2 \doteq S_{B+}^2$ and

$$S_{AB}^2 = S_{A+B+}^2 = \frac{1}{2}(S_A^2 + S_B^2) + \frac{1}{4}\Delta_{AB}^2 \quad . \quad (A.9)$$

Combining this result (A.9) with that for the covariance (A.5) gives

$$R_{1,AB} = \frac{\text{cov}_A + \text{cov}_B + \frac{1}{2}\Delta_{AB}^2}{S_A^2 + S_B^2 + \frac{1}{2}\Delta_{AB}^2}$$

$$= \frac{r_A S_A^2 + r_B S_B^2 + \frac{1}{2}\Delta_{AB}^2}{S_A^2 + S_B^2 + \frac{1}{2}\Delta_{AB}^2} \quad (A.10)$$

because $r_A = cov_{AA+}/S_A S_{A+} \triangleq cov_A/S_A^2$. Adding and subtracting the mean $\bar{r} = (r_A + r_B)/2$, and dropping the subscripts on Δ , gives

$$R_{1,AB} = \bar{r} + \frac{2r_A S_A^2 + 2r_B S_B^2 + \Delta^2 - \frac{1}{2}(r_A + r_B)(2S_A^2 + 2S_B^2 + \Delta^2)}{2(S_A^2 + S_B^2) + \Delta^2}$$

$$= \bar{r} + \frac{r_A(2S_A^2 - S_A^2 - S_B^2) + r_B(2S_B^2 - S_A^2 - S_B^2) + \Delta^2(1 - \bar{r})}{2(S_A^2 + S_B^2) + \Delta^2}$$

$$= \bar{r} + \frac{1 - \bar{r}}{1 + 2(S_A^2 + S_B^2)/\Delta^2} + \frac{(r_A - r_B)(S_A^2 - S_B^2)}{\Delta^2 + 2(S_A^2 + S_B^2)} \qquad \text{(A.11)}$$

This is Eq. (5.1), which is examined farther in Section 5.

TABLE 5 SERIAL CORRELATIONS FOR 6 Hr. Lag, 0-27 Kilometers

	CALENDAR INTERVAL												
	0	1	2	3	4	5	6	7	8	9	10	11	12
0000 - 0600													
Scalar Speed	.95	.91	.90	.94	.82	.88	.72	.87	.91	.94	.96	.92	.96
Scalar Momentum	.95	.90	.84	.91	.78	.84	.75	.67	.90	.92	.94	.93	.92
Vector Speed	.96	.90	.92	.92	.92	.90	.91	.90	.97	.93	.96	.91	.96
Vector Momentum	.94	.91	.85	.90	.89	.88	.87	.61	.94	.89	.94	.93	.92
0600 - 1200													
Scalar Speed	.94	.95	.90	.95	.90	.95	.85	.90	.93	.94	.98	.93	.97
Scalar Momentum	.93	.94	.86	.94	.89	.92	.85	.85	.82	.93	.97	.93	.97
Vector Speed	.95	.95	.93	.97	.93	.92	.95	.95	.97	.97	.97	.91	.97
Vector Momentum	.92	.94	.86	.97	.93	.94	.87	.86	.91	.95	.97	.92	.98
1200 - 1800													
Scalar Speed	.95	.94	.94	.95	.90	.90	.89	.86	.87	.94	.98	.93	.94
Scalar Momentum	.94	.95	.96	.95	.87	.82	.90	.83	.88	.91	.97	.90	.95
Vector Speed	.95	.94	.95	.97	.96	.84	.96	.90	.96	.95	.98	.88	.96
Vector Momentum	.94	.94	.95	.97	.93	.84	.90	.82	.93	.92	.97	.89	.95
1800 - 0000													
Scalar Speed	.96	.93	.90	.93	.96	.91	.84	.84	.82	.86	.97	.93	.93
Scalar Momentum	.95	.92	.83	.92	.92	.90	.81	.72	.77	.80	.96	.93	.92
Vector Speed	.96	.92	.94	.97	.95	.75	.95	.93	.95	.89	.98	.92	.94
Vector Momentum	.96	.93	.85	.96	.90	.90	.89	.79	.89	.82	.97	.93	.92
All Hours													
Scalar Speed	.95	.92	.91	.94	.88	.90	.81	.86	.87	.92	.97	.92	.95
Scalar Momentum	.94	.91	.87	.93	.85	.85	.81	.76	.84	.89	.96	.92	.94
Vector Speed	.95	.92	.93	.96	.93	.85	.94	.91	.96	.93	.97	.90	.96
Vector Momentum	.94	.92	.87	.95	.91	.88	.88	.75	.92	.90	.96	.91	.94

TABLE 6 SERIAL CORRELATIONS FOR 6 Hr. Lag, 0-6 Kilometers

	CALENDAR INTERVAL												
	0	1	2	3	4	5	6	7	8	9	10	11	12
0000 - 0600													
Scalar Speed	.92	.91	.76	.85	.92	.56	.78	.65	.81	.84	.91	.88	.86
Scalar Momentum	.92	.89	.76	.85	.93	.57	.77	.60	.80	.82	.89	.88	.86
Vector Speed	.92	.92	.79	.88	.95	.74	.83	.64	.83	.82	.93	.89	.86
Vector Momentum	.92	.91	.79	.87	.96	.78	.84	.58	.84	.78	.91	.89	.86
0600 - 1200													
Scalar Speed	.89	.93	.82	.86	.94	.75	.85	.84	.48	.86	.97	.87	.97
Scalar Momentum	.90	.93	.81	.86	.94	.74	.86	.82	.49	.85	.96	.88	.97
Vector Speed	.90	.92	.82	.90	.94	.82	.89	.84	.67	.86	.97	.88	.97
Vector Momentum	.90	.91	.81	.89	.94	.81	.92	.82	.70	.83	.97	.89	.97
1200 - 1800													
Scalar Speed	.92	.91	.94	.86	.84	.62	.86	.87	.87	.88	.93	.81	.92
Scalar Momentum	.93	.90	.94	.85	.86	.62	.88	.85	.89	.88	.92	.82	.92
Vector Speed	.92	.90	.93	.87	.84	.74	.90	.87	.90	.90	.93	.85	.91
Vector Momentum	.92	.89	.93	.86	.88	.77	.90	.85	.90	.89	.91	.86	.92
1800 - 0000													
Scalar Speed	.95	.91	.73	.84	.86	.83	.82	.73	.73	.71	.93	.90	.93
Scalar Momentum	.94	.91	.74	.83	.86	.81	.82	.72	.73	.76	.92	.91	.92
Vector Speed	.95	.92	.74	.88	.89	.85	.87	.70	.81	.59	.95	.91	.92
Vector Momentum	.94	.92	.75	.87	.89	.83	.87	.64	.82	.64	.94	.91	.92
All Hours													
Scalar Speed	.92	.91	.81	.85	.87	.69	.81	.76	.72	.82	.93	.86	.92
Scalar Momentum	.92	.90	.81	.84	.88	.67	.81	.74	.73	.82	.92	.87	.92
Vector Speed	.92	.91	.81	.88	.90	.78	.87	.75	.80	.79	.94	.88	.91
Vector Momentum	.92	.91	.81	.87	.91	.80	.87	.71	.81	.78	.93	.89	.91

TABLE 7 SERIAL CORRELATIONS FOR 6 Hr. Lag, 7-13 Kilometers

	CALENDAR INTERVAL												
	0	1	2	3	4	5	6	7	8	9	10	11	12
0000 - 0600													
Scalar Speed	.94	.87	.89	.93	.81	.90	.60	.88	.91	.90	.93	.92	.94
Scalar Momentum	.94	.86	.89	.92	.80	.90	.65	.84	.90	.90	.92	.92	.93
Vector Speed	.94	.87	.89	.93	.83	.92	.55	.86	.92	.87	.94	.92	.94
Vector Momentum	.95	.85	.88	.92	.83	.92	.61	.79	.92	.88	.93	.92	.94
0600 - 1200													
Scalar Speed	.93	.93	.88	.92	.91	.94	.78	.93	.92	.89	.96	.93	.95
Scalar Momentum	.93	.92	.89	.93	.90	.93	.80	.92	.91	.89	.95	.93	.95
Vector Speed	.93	.92	.88	.94	.89	.95	.74	.94	.91	.91	.95	.93	.96
Vector Momentum	.93	.91	.89	.94	.88	.94	.75	.93	.91	.91	.95	.93	.95
1200 - 1800													
Scalar Speed	.93	.94	.89	.93	.93	.92	.76	.73	.90	.91	.96	.91	.95
Scalar Momentum	.92	.94	.88	.94	.93	.91	.75	.68	.89	.90	.96	.90	.94
Vector Speed	.93	.94	.89	.94	.94	.93	.72	.73	.89	.88	.97	.91	.95
Vector Momentum	.93	.94	.88	.94	.93	.92	.70	.69	.89	.87	.97	.91	.94
1800 - 0000													
Scalar Speed	.93	.92	.85	.94	.97	.87	.80	.81	.86	.79	.96	.94	.88
Scalar Momentum	.92	.91	.84	.94	.96	.87	.78	.75	.85	.79	.96	.94	.88
Vector Speed	.92	.92	.85	.94	.97	.88	.83	.82	.86	.83	.96	.94	.88
Vector Momentum	.91	.91	.84	.94	.96	.88	.83	.76	.83	.84	.96	.94	.88
All Hours													
Scalar Speed	.93	.90	.88	.93	.89	.89	.72	.83	.89	.87	.95	.92	.93
Scalar Momentum	.93	.89	.87	.93	.88	.88	.72	.79	.89	.86	.95	.92	.92
Vector Speed	.93	.90	.88	.93	.90	.90	.69	.83	.89	.87	.95	.92	.93
Vector Momentum	.93	.89	.87	.93	.89	.90	.69	.78	.89	.87	.95	.92	.92

TABLE 8 SERIAL CORRELATIONS FOR 6 Hr. Lag, 14-20 Kilometers

	CALENDAR INTERVAL												
	0	1	2	3	4	5	6	7	8	9	10	11	12
0000 - 0600													
Scalar Speed	.93	.81	.94	.94	.97	.83	.77	.88	.88	.86	.96	.88	.97
Scalar Momentum	.94	.83	.93	.94	.97	.88	.81	.88	.90	.87	.97	.88	.97
Vector Speed	.93	.80	.95	.95	.96	.85	.92	.91	.93	.92	.97	.86	.97
Vector Momentum	.94	.83	.94	.94	.97	.90	.94	.95	.94	.90	.98	.87	.97
0600 - 1200													
Scalar Speed	.94	.87	.94	.93	.87	.81	.88	.88	.90	.82	.98	.89	.96
Scalar Momentum	.94	.87	.94	.93	.91	.86	.87	.89	.92	.84	.98	.91	.95
Vector Speed	.94	.89	.94	.92	.86	.87	.95	.94	.91	.90	.99	.92	.96
Vector Momentum	.94	.88	.94	.93	.90	.91	.94	.93	.94	.89	.98	.92	.95
1200 - 1800													
Scalar Speed	.92	.93	.95	.95	.89	.86	.86	.89	.92	.86	.97	.93	.96
Scalar Momentum	.93	.93	.95	.95	.92	.89	.84	.88	.94	.89	.97	.92	.96
Vector Speed	.92	.93	.95	.95	.93	.90	.93	.92	.95	.84	.97	.94	.96
Vector Momentum	.93	.93	.96	.96	.95	.91	.92	.92	.96	.87	.97	.93	.96
1800 - 0000													
Scalar Speed	.92	.93	.94	.94	.91	.85	.85	.86	.82	.90	.98	.92	.97
Scalar Momentum	.93	.93	.94	.94	.92	.88	.85	.88	.81	.90	.98	.91	.96
Vector Speed	.91	.92	.94	.94	.92	.92	.92	.90	.92	.92	.98	.92	.97
Vector Momentum	.93	.93	.95	.95	.94	.95	.92	.91	.92	.91	.98	.91	.96
All Hours													
Scalar Speed	.93	.87	.93	.94	.90	.83	.77	.84	.83	.85	.97	.90	.96
Scalar Momentum	.93	.88	.93	.94	.93	.87	.81	.86	.87	.87	.97	.91	.96
Vector Speed	.92	.87	.93	.93	.91	.87	.88	.89	.91	.89	.98	.91	.96
Vector Momentum	.93	.89	.94	.94	.94	.91	.90	.91	.93	.88	.98	.91	.96

TABLE 9 SERIAL CORRELATIONS FOR 6 Hr. Lag, 21-27 Kilometers

	CALENDAR INTERVAL												
	0	1	2	3	4	5	6	7	8	9	10	11	12
0000 - 0600													
Scalar Speed	.89	.97	.89	.84	.86	.92	.75	.78	.86	.96	.78	.85	.82
Scalar Momentum	.86	.96	.86	.79	.88	.92	.70	.77	.87	.96	.74	.77	.87
Vector Speed	.90	.97	.84	.91	.93	.91	.80	.78	.89	.96	.83	.92	.68
Vector Momentum	.89	.96	.82	.89	.95	.91	.79	.77	.89	.95	.70	.87	.76
0600 - 1200													
Scalar Speed	.82	.91	.83	.81	.92	.91	.87	.82	.89	.96	.72	.87	.88
Scalar Momentum	.77	.90	.80	.75	.92	.94	.89	.82	.89	.94	.70	.84	.90
Vector Speed	.85	.92	.90	.91	.91	.92	.86	.83	.90	.95	.83	.92	.82
Vector Momentum	.84	.91	.85	.82	.92	.94	.90	.83	.90	.93	.77	.90	.82
1200 - 1800													
Scalar Speed	.83	.90	.85	.70	.83	.92	.84	.79	.88	.97	.63	.89	.69
Scalar Momentum	.81	.90	.81	.70	.87	.92	.82	.82	.90	.96	.64	.88	.72
Vector Speed	.87	.91	.88	.87	.91	.92	.84	.79	.89	.97	.71	.90	.86
Vector Momentum	.86	.90	.84	.86	.93	.92	.83	.81	.90	.96	.69	.90	.87
1800 - 0000													
Scalar Speed	.90	.93	.84	.81	.91	.91	.88	.83	.86	.95	.71	.90	.77
Scalar Momentum	.89	.93	.76	.75	.90	.90	.83	.81	.86	.93	.63	.86	.77
Vector Speed	.90	.91	.86	.91	.94	.91	.87	.82	.87	.94	.80	.92	.66
Vector Momentum	.86	.92	.80	.85	.95	.90	.84	.79	.87	.93	.77	.90	.69
All Hours													
Scalar Speed	.84	.92	.83	.79	.87	.90	.76	.74	.82	.95	.67	.87	.78
Scalar Momentum	.81	.91	.79	.74	.88	.89	.70	.70	.82	.93	.63	.83	.81
Vector Speed	.86	.92	.86	.89	.92	.90	.78	.74	.84	.95	.79	.91	.73
Vector Momentum	.84	.91	.82	.84	.93	.89	.73	.71	.84	.93	.72	.88	.77

MATHEMATICAL WIND PROFILES

PART IV (Final Report)

SUMMARY

Properties of augmented complex Fourier polynomials, developed to represent the complete vertical profile of horizontal wind vectors, are summarized from three previous reports. Procedures for predicting a future vector profile from present and past profiles are developed in detail, for application to winds at Cape Kennedy, Florida.

1. INTRODUCTION (IV)

Description and prediction of vector wind profiles has been an unsolved meteorological problem of increasing importance for many years. No standard procedure for the mathematical description of a profile has yet emerged, despite many attacks on the problem. Forecasting of an entire profile has been even less successful, and most forecast procedures are for the winds at individual levels, separately, rather than for the whole profile.

Mathematically, a vector wind profile is a continuous function of three variables: height, and two horizontal components of the wind at each height. (The vertical component of wind is neglected because it is some two orders of magnitude smaller than the resultant of the two horizontal components.) The quest for a suitable representation of such a 3-dimensional function, which could be used in predicting the wind profile at some future time, or over a specified place for which data are not immediately available, was undertaken under Contract NAS 8-5380 with the Marshall Space Flight Center, NASA, Huntsville, Alabama.

In this fourth and final report on that contract, first the development and testing of a method of representing wind profiles is summarized; details were given in three previous reports[1,2,3], designated hereafter as MWP-I, MWP-II, MWP-III. Most of this report is devoted to prediction schemes for wind profiles, as described mathematically by augmented complex Fourier polynomials.

Such trigonometric polynomials, from the detailed studies reported in MWP-I and MWP-II, seem appropriate for describing vector wind profiles. They are direct extensions, to complex variables, of standard Fourier polynomials, which are complex-valued functions of real variables and have been used in many branches of applied mathematics for more than a century. Apparently complex Fourier polynomials have not been used hitherto for geophysical problems such as vector wind profiles.

For adequate representation of vector wind profiles, the Fourier polynomials of a complex argument must be augmented, as explained later, by terms describing any linear trends in each of the wind components.

2. BASIC FORMULAS

In the application of this Fourier representation, the eastward (x) and northward (y) components of the wind at any height h are combined into a complex variable $z = x + iy$. Observations on z at $n + 1$ equally-spaced heights, $h = 0, 1, 2, \ldots, n$, are used to compute $2(n + 2)$ complex coefficients. They permit the estimation of the wind vector at any height as the sum of $n + 2$ complex quantities: a complex constant c_z, the product of a complex coefficient d_{oo} and the height h, and n products of complex coefficients d_j and complex numbers:

$$z(h) = c_z + d_{oo} h + \sum_{j=0}^{n} d_j \exp(i \lambda j h), \qquad (2.1)$$

where $\lambda = 2\pi/\nu$ and $\nu = n + 1$. As explained later, $d_o = 0$.

The first two terms represent the linear trend, which can be removed in any of several ways. A plane can be fitted to the vector wind profile by ordinary least squares or by some modification so that its distances from the two endpoints of the profile are equal. Alternatively, trend can be removed from each component separately. Extensive research (some reported in MWP-I and discussed in MWP-II) suggests that simple least squares linear regression on each component separately is adequate. At each level h for which wind data are available, the trend lines are

$$x = u + c_x + a_{oo} h \quad , \qquad y = v + c_y + b_{oo} h \quad . \qquad (2.2)$$

The linear constants $c_z = c_x + i c_y$ and coefficients $d_{oo} = a_{oo} + i b_{oo}$ are estimated as

$$a_{oo} = (\nu \Sigma x_h h - \Sigma x_h \Sigma h)/\nu \sigma_h^2$$

$$b_{oo} = (\nu \Sigma y_h h - \Sigma y_h \Sigma h)/\nu \sigma_h^2$$

$$c_x = \bar{x} - a_{oo} \bar{h}, \quad c_y = \bar{y} - b_{oo} \bar{h} \tag{2.3}$$

where $\sigma_h^2 = n(n+2)/12$ because $h = 0, 1, 2, \ldots, n$ and hence $\bar{h} = n/2 = (\nu - 1)/2$.

The rest of the Fourier representation is in terms of the departures u and v from these trend lines. They are formed into a new variable $w = u + iv$, which has a mean of zero. Ordinary Fourier representation has $n + 1$ terms, of which the first is a constant, representing the mean. The trend removal in effect replaces this constant term by two terms, so that the complete representation requires only one additional term than the ordinary procedure.

The complex Fourier coefficients $d_j = a_j + i b_j$ are estimated by the method of least squares as

$$a_j = \frac{1}{\nu} \sum_{h=0}^{n} \left[u_h \cos(\lambda j h) + v_h \sin(\lambda j h) \right]$$

$$b_j = \frac{1}{\nu} \sum_{h=0}^{n} \left[v_h \cos(\lambda j h) - u_h \sin(\lambda j h) \right] . \tag{2.4}$$

For $j = 0$, $a_o = b_o = 0$ because $\bar{u} = \bar{v} = 0$. The variance of w is given by

$$s_w^2 = \frac{1}{\nu} \left[\sum_{h=0}^{n} u_h^2 + \sum_{h=0}^{n} v_h^2 \right] = \sum_{j=0}^{n} A_j^2 , \tag{2.5}$$

where $A_j^2 = a_j^2 + b_j^2$. The variance of the original variable z is computed as

$$s_z^2 = s_x^2 + s_y^2 \qquad (2.6)$$

where $s_x^2 = \frac{1}{\nu} \sum_{h=0}^{n} (x_h - \bar{x})^2$ and $s_y^2 = \frac{1}{\nu} \sum_{h=0}^{n} (y_h - \bar{y})^2$.

A FORTRAN IV program for computing the Fourier coefficients, and calculating the variances, is given in Appendix A.

At each altitude h for which wind data are given originally, the sum of the $n + 2$ complex terms provides a computed wind which agrees precisely with the observed winds. At any intermediate level, the estimated wind is a function of all the observed winds, not just an interpolation between values at the two adjacent levels of observation. Hence this procedure can be used to estimate winds at any level in a sounding.

The method is not restricted to the representation of wind vectors. It can be applied equally well to momentum, which is a vector obtained by multiplying the wind vector by the atmospheric density, a scalar quantity. The method can be applied also to the positions of a balloon, observed at fixed time intervals, to give a position-time representation. A suitable transformation, using the observed rate-of-rise of the balloon, will convert this into a position height representation, from which wind speeds can be obtained by differentiation. Although this is a very promising prospect for more accurate reduction of upper wind observations than techniques in current use, it was not developed to the point of application because it was outside the primary purpose of the research.

3. VARIANCE REDUCTION

The basic augmented complex Fourier representation of a vector wind profile (2.1) has a complex constant and $n + 1$ complex coefficients, derived from wind observations at $n + 1$ equally-spaced levels. In most cases, many of these coefficients are very small, indicating that only a few of the terms are strongly significant. The significance of each term is indicated by its contribution to the reduction in the variance of the difference between the estimated and actual winds.

Linear trends account for a substantial portion of the variance in the basic wind observations, because of the general increase in wind speed with height to a level of extreme wind, usually somewhat below the tropopause, and a decrease of wind with height above this level. The percent reduction in variance arising from removal of trend is

$$100\, s_L^2 = 100\,(1 - s_w^2 / s_z^2)\,. \qquad (3.1)$$

The percent of the total variance "explained" by the jth Fourier term is

$$100\, s_j^2 = 100\, A_j^2 / s_z^2\,. \qquad (3.2)$$

The partitioning of the total variance (2.5) into ν parts in terms of the Fourier coefficients is a consequence of the Parseval identity. Orthogonality of the d_j implies independence and allows for meaningful partitioning of the total variance into components of variance for each term; the d_j are almost orthogonal (MWP-I, App. B). The percent variance explained by each term is a measure of its importance in the Fourier representation.

Extensive research (reported in MWP-I and MWP-II) on wind soundings at 1-km intervals between 2 and 25 km, giving data at 24 levels, showed that 82 to 95 percent of the total variance would be removed by the linear terms and only four Fourier terms, generally those for $j = 1, 2, 22$, and 23; in some cases other terms, notably 21, were slightly more important than 22. For other soundings, with data at a greater number of levels, generally the first two and the last two terms provided the greatest reduction in variance.

Thus for some purposes an adequate representation of a sounding can be obtained by computing all the $n + 2$ complex coefficients, ranging them according to the percent of variance represented (or explained) by each, and then using only the half-dozen best (highest ranking) terms for the representation. This approach, while not as precise as the use of all terms, may be needed for certain purposes, especially prediction, as discussed later.

4. PREDICTION

Augmented complex Fourier polynomials, which describe a complete wind profile by the use of coefficients computed from observations at equally-spaced height intervals, can be the basis of a wind prediction method. The various coefficients, rather than the winds themselves, are predicted by some regression procedure, then combined (2.1) to yield the predicted winds at any level of interest.

A suitable set of prediction equations is developed, conventionally in four steps:

a. The components and functional form of a prediction model are established.

b. The unknown parameters of the model are estimated from a set of known data.

c. These estimated parameters, and the adopted function, are used to make independent predictions.

d. The predicted values are compared with actual ones, and the degree of correspondence computed.

For purposes of this discussion, the $n + 2$ complex coefficients c_z, d_{oo}, and d_j are separated into the real and imaginary parts, giving $2n + 4$ coefficients. These are denoted as g_k, using the subscript k to identify the particular coefficient. Then g_{ki} is the kth coefficient in the Fourier representation of the ith wind sounding in some series of soundings, generally at the same place in the same season.

The general problem is to represent the coefficients $g_{k,i+1}$ of the $i + 1$st sounding as functions of the coefficients g_{ki} of some or all of the preceding I soundings. This requires the assumption of some functional form for the dependence of the $i + 1$st coefficients on the preceding ones:

$$g_{k,i+1} = f(g_{ki}) \ , \quad i = 1, 2, \ldots, I . \tag{4.1}$$

Several prediction models are possible, using different functional forms. The function f may be linear, quadratic, exponential or anything else. Thus simple linear regression would be

$$g_{k,i+1} = B_{0k} + B_{1k} g_{k,i} + B_{2k} g_{k,i-1} \cdots \tag{4.2}$$

Polynomial regression would be

$$g_{k,i+1} = B_{0k} + \sum_{j=1}^{p} \sum_{\alpha=1}^{m} B_{jk\alpha} \, g_j^{\alpha} \tag{4.3}$$

which reduces to the linear form when $m = 1$.

In terms of the Fourier representation of a sounding, $g_{k,i+1}$ could be assumed to be related only to the same coefficient, g_{ki}, on the

previous sounding, to all such preceding coefficients, to all the coefficients on the preceding sounding and no others, or any other combination. No valid basis could be found for the assumption that the kth coefficient on one sounding is solely a function of the corresponding coefficient on the preceding sounding. This would imply that the complete Fourier form is fixed from observation to observation. Hence each coefficient was assumed to depend on all the coefficients of the preceding sounding, and perhaps on those of a few previous soundings.

That prediction would be more accurate in terms of a momentum vector than a velocity vector was suggested by elementary considerations of the conservation of energy in the atmosphere. Physical reasoning, and experience, also suggested that the diurnal cycle might be important in wind prediction, especially in the subtropics, and that the wind 24 hour prior to the prediction time might be of some use in prediction, although the latest available previous wind, generally 6 hours before prediction time, would still be the most useful.

To test these suggestions, serial correlation coefficients were computed (MWP-III) for one year (1962) of wind observations at 6 hour intervals over Cape Kennedy. Any missing data had been previously interpolated to complete the set of data. Correlations were computed for integrated atmospheric zones for four basic wind representations (scalar wind, vector wind, scalar momentum, vector momentum) at lags of 1, 2, 3, and 4, time intervals of 6 hours each.

Vector speed showed the highest serial correlation, and the correlation for lag 1 (6 hour) was much higher than for the other lags, including 24 hours - which was not as good as the 12 hour lag. Thus the prediction model was developed for coefficients of the Fourier representation of vector wind, using the corresponding representation of the wind 6 hours earlier.

5. FORMULATION

The regression model for predicting the Fourier coefficients of a vector wind sounding from those of the preceding sounding can be developed most readily in matrix notation. The notation to be used is generally standard, but is summarized here for convenience. Matrices are denoted by capital letters, their elements by lower case letters. The first subscript on an element always refers to the row, the second to the column position. Rows are numbered downward, columns from left to right. For example, in matrix $X_{n \times p}$ of n rows and p columns, x_{ik} is in the ith row and kth column.

The transpose of a matrix, denoted by the superscript T, has rows and columns interchanged: in X^T, the element x_{ki} in the kth row and ith

column is element x_{ik} (ith row, kth column) of the original matrix X. Two matrices can be multiplied only if the first has as many columns as the second has rows. The product of $U(A \times B)$ and $V(B \times C)$ is $W(A \times C)$. The transpose of a matrix product is the reversed product of the transposes:

$$(UV)^T = W^T = V^T U^T .$$

Specifically, $U^T U$ is $(B \times B)$ while $U U^T$ is $(A \times A)$. A vector is a matrix with only one column; its transpose has only one row.

The principal diagonal of a matrix is the sequence of elements x_{ii}. The identity matrix I has all elements on the principal diagonal of unity and zeros elsewhere. The inverse of a matrix, denoted by a -1 superscript, is defined as the matrix such that

$$X^{-1} X = X X^{-1} = I .$$

A matrix has a unique inverse only when it is of full rank, or non-singular, that is, when no row or column is a linear combination of any other row or column.

The adopted linear regression model is Eq. (4.2) for $g_{k,i+1}$ as a linear function of all the $2n + 2$ coefficients $g_{k,i}$ of the Fourier representation of the previous (ith) sounding:

$$g_{k,i+1} = \beta_{k,0} + \beta_{k,1} g_{1,i} + \beta_{k,2} g_{2,i} + \cdots + \beta_{k,p} g_{p,i} \qquad (5.1)$$

for $k = 1, 2, \ldots, 2n + 2$, and for any i. The $2n + 2 = p$ equations are, in matrix form, $G_{i+1} = B_o + B_k G_i^o$ or

$$\begin{bmatrix} g_{1,i+1} \\ g_{2,i+1} \\ \cdots \\ g_{p,i+1} \end{bmatrix} = \begin{bmatrix} \beta_{1,0} \\ \beta_{2,0} \\ \cdots \\ \beta_{p,0} \end{bmatrix} + \begin{bmatrix} \beta_{1,1} & \cdots & \beta_{1,p} \\ \beta_{2,1} & \cdots & \beta_{2,p} \\ \cdots & \cdots & \cdots \\ \beta_{p,1} & \cdots & \beta_{p,p} \end{bmatrix} \begin{bmatrix} g_{1,i} \\ g_{2,i} \\ \cdots \\ g_{p,i} \end{bmatrix} .$$

The awkward addition of matrices can be avoided by adding a dummy variable $g_{0,i} \equiv 1$ to G_i^o, calling the new vector of $p + 1$ elements G_i, and combining B_o and B_k into a $p \times (p + 1)$ matrix B:

$$\begin{bmatrix} g_{1,i+1} \\ g_{2,i+1} \\ \cdots \\ g_{p,i+1} \end{bmatrix} = \begin{bmatrix} \beta_{1,0} & \beta_{1,1} & \cdots & \beta_{1,p} \\ \beta_{2,0} & \beta_{2,1} & \cdots & \beta_{2,p} \\ \cdots & \cdots & \cdots & \cdots \\ \beta_{p,0} & \beta_{p,1} & \cdots & \beta_{p,p} \end{bmatrix} \begin{bmatrix} g_{0,i} \\ g_{1,i} \\ \cdots \\ g_{p,i} \end{bmatrix} \qquad (5.2)$$

or, simply, $G_{i+1} = B\, G_i$.

Polynomial matrix regression, if desired, would be

$$G_{i+1} = B\, G_i + B^{[2]}\, Q^{[2]} + \cdots + B^{[m]}\, Q^{[m]} \qquad (5.3)$$

where $Q^{[2]}$ is the principal diagonal of $G^T G$ and $Q^{[m]}$ is the principal diagonal of $(Q^{[m-1]})^T G$, each formed into a p-element vector.

In many respects, the m-degree polynomial would be preferable to simple linear matrix regression. Available data and estimation problems suggest, however, that the linear model is sufficiently complicated for the first attempt at a regression model. It will not give a better estimate than the polynomial, but with a relatively small amount of data the linear model will have somewhat narrower confidence intervals around its estimates than would the polynomial.

6. COMPUTATION

Estimation of the elements of the prediction matrix B (5.2) for a specific location, atmospheric height interval, and month or season requires a set of appropriate wind soundings. The augmented complex Fourier representation (2.1) must first be obtained for each sounding, so that it is represented by $2n + 2 = p$ coefficients $g_{1,i} \ldots, g_{p,i}$. For computational convenience, the first index is increased by one, so that the terms are $g_{2,i} \ldots, g_{p+1,i}$, and a dummy variable $g_{1,i} \equiv 1$ is introduced.

In applying the proposed wind vector profile forecasting procedure for Cape Kennedy, one year (1962) of serially-completed soundings at 6 hour intervals was available. To investigate serial correlation (MWP-III), they had been divided into 13 consecutive time periods of 28 days or 112 soundings each. Each sounding provided wind data at 1 km intervals from the surface to 27 km, but to eliminate the problem of winds in the surface layer (MWP-I), data from only 3 to 26 km were used. Thus each sounding had wind data from $n + 1 = 24$ levels, represented by $2n + 2 = p = 50$ coefficients preceded by the dummy coefficient, or 51 in all. To provide for the continuity between the end of one time interval and the start of the next, each 28-day interval was preceded by the four soundings of the previous one, making a total of $112 + 4 = 116$ soundings. Thus for each time period, the $q + 1 = 116$ representations, each of 51 coefficients, form a 51 x 116 element matrix.

In this matrix, the ith column contains a one followed by all the coefficients of the ith sounding. Discarding the last column gives a matrix G, and discarding the first column gives another matrix S; each is $(p + 1) \times q$, where $p = 2n + 2$ and n is the number of equally-spaced levels for which wind data were obtained, while q is one less than the total number of soundings, made at equal time intervals. The ith column of S is the same as the $i + 1$st column of G. Hence, according to the regression model, $S_i = B G_i$. In terms of these two matrices, the least squares, and also maximum likelihood, estimate of B is

$$\hat{B} = (G G^T)^{-1} G S^T \qquad (6.1)$$

The form of this estimating equation differs slightly from that often used in multiple linear regression because, for this matrix linear regression, the columns, and not the rows, of G and S represent individual observations -- in this case, Fourier representations of vertical wind soundings.

The major problem in estimating B from (6.1) is in inverting the matrix $G G^T$. A unique solution for B will be obtained only if the matrix is non-singular, and if a computer with sufficient capacity is available for the inversion. Most matrix inversion programs contain a singularity check before starting the actual inversion procedure.

The possibility that $G G^T$ is singular can be lessened by eliminating some of the coefficients in G, and thus reducing the number of rows. In Fourier representations of vector wind profiles, many of the coefficients contribute negligible reductions in the error variance (MWP-II). Equality of the coefficients g_{ki} for two or more k and for every i may cause singularity, and the most likely cause for equality is that the coefficients are actually zero. Thus any program in which soundings are represented by a large number of coefficients - as would be the case in applying the complex Fourier

representation to detailed wind soundings with data points every 10 or 20 meters - should include an option for selecting only the more significant coefficients (MWP-I, MWP-II) for use in a prediction model.

Matrix inversion procedures are extremely sensitive to round-off error. Even though computer storage can be increased markedly by addition of storage units and by programming ingenuity, any computer must have some upper limit to the size of matrix that can be inverted. For example, the IBM 7094 of the Lockheed-California Company can invert a 2000 by 2000 matrix of 8-digit elements. For wind prediction by augmented complex Fourier polynomials, this limits the number of intervals in the original soundings to 997, equivalent to wind observations every 20 meters through a 20 km thickness, or every 50 meters through 50 km. But if each sounding is represented by only half of the available coefficients, twice as many levels can be evaluated in the first place.

7. CONCLUSIONS (IV)

The research program summarized here and in the three previous reports was undertaken to develop a method for describing and forecasting complete vector wind profiles. Of all the methods of representation that were considered, Fourier polynomials of a complex variable seemed best suited. However, the characteristics of vector wind profiles required that such polynomials be augmented by linear terms.

Most of the investigation was devoted to developing computation procedures for augmented complex Fourier polynomials and to studying their properties: prediction would be futile if the Fourier representations were not stable, but changed markedly from sounding to sounding, or depended strongly on the particular height interval over which the representation was made.

After the properties were found to be generally satisfactory for description, prediction methods and procedures were developed. Vector velocity was found to have greater time consistency than scalar speed or scalar or vector momentum, and 6 hour lag correlations were much greater than for longer lags. Consequently, the first prediction model was for regression of a vector velocity representation on the corresponding representation 6 hours earlier.

Computer programming of the regression model is simple, but the estimation of the regression matrix encountered several difficulties, primarily connected with matrix inversion. Several successive modifications of the basic program did not yield a usable solution before termination of the study; perhaps an additional 100 man-hours would suffice to obtain a working program with which regression matrices could be computed. Programming requirements are discussed in Appendix B.

In summary, a method for representing the complete vector wind profile was developed and some of its properties ascertained. Its usefulness in a prediction model could not be determined, however, because unexpected computer programming difficulties prevented the obtaining of the required matrix of regression coefficients. Hence no actual predictions were made for testing goodness of fit.

Even if the regression procedure proves unsatisfactory for short-range prediction of vector wind profiles, the augmented complex Fourier polynomials, developed for such intended use, offer one potentially valuable application. Such a polynomial can be used to represent the time-position hodograph of a rising (or falling) balloon or other radar target, transformed into a position-height function, and differentiated to provide wind velocities without the smoothing and interpolation inherent in present methods. Further research and development of this application seems desirable for better measurement of wind.

APPENDIX IV-A

COMPLEX FOURIER ANALYSIS PROGRAM

The complete FORTRAN IV computer program for determining and evaluating the coefficients of the augmented complex Fourier polynomial describing a single vector wind profile is given in Table A-1. The first part of the program finds and removes the linear trend lines in the two components x and y, to form the new variables u and v, computes the complex Fourier coefficients for each term and determines the percent of total variance that is represented (or "explained") by each complex term. Input to the program is a single sounding: values of x and y at n + 1 equally-spaced points, i.e. atmospheric heights. These values may be wind speed, or merely position coordinates; if wind speed, they may be multiplied by the corresponding density (RHO) to make momentum the input.

In the second part, which is optional, the Fourier coefficients are used to compute u - v profiles using only one, only two, only three, etc. to all n terms, selected in consecutive cumulative order starting with the first coefficients (j = 1) to the last (j = n).

In the program output format, the first line gives all quantities related to the trend removal in the original x and y :

SX2: variance of x

SY2: variance of y

A1: constant term for x (denoted as c_x in 2.3)

B1: linear coefficinet for x (denoted as a_{00} in 2.3)

A2: constant term for y (denoted as c_y in 2.3)

B2: linear coefficient for y (denoted as b_{00} in 2.3)

The second line contains headings for the original and reduced soundings, the Fourier coefficients, and percent variance explained. The index H refers to the height level, reindex to run from 0 to N ; no provision is made for reverting to actual heights, as specified in the original input. The index J identifies the complex Fourier coefficients; coefficients for J = 0 are 0 because of trend line removal.

X(H) , Y(H): original values of input sounding

U(H) , V(H): residual values of sounding after trend removal

R(J) , C(J): real and imaginary parts of Jth complex coefficient

S(J) : fraction of reduced variance, in terms of u and v , represented ("explained") by Jth Fourier term:

$$S(J) = A_j^2/V \text{ , where } A_j^2 = R^2(J) + C^2(J) \text{ , } V = \Sigma_J A_j^2$$

P(J) : percent of total variance, in terms of x and y, represented by Jth Fourier term: P(J) = 100 Q S(J), where Q = V/T and T = SX2 + SY2.

The next-to-last line gives a check on the zero-mean variables u and v, and the reduced variance $V = \Sigma A_j^2$.

SUM U SUB H and SUM V SUB H are summations of u and v over H. The last line gives R = 100 (1 -Q), which is the percent of the total variance represented by the linear terms.

A sample of the program output is given in Table A-2 for the sounding of 1000Z on 8 June 1956 over Montgomery, Alabama.

TABLE A-1

```
$IBFTC FOUR1
CWIND
C       LEAST SQUARES FIT OF A FOURIER SERIES
        DIMENSION X(200),Y(200),U(200),V(200),R(200),C(200),A(200),S(200)
       1 ,RR(200),CC(200),UH(600),VH(600)      ,XX(200),YY(200)
        DIMENSION  P(200), RHO(200)
 1      FORMAT (7I3)
 2      FORMAT(7F10.4)
 3      FORMAT (1H  7F16.5)
 4      FORMAT(1H  I3, 8F14.5)
  5     FORMAT(1H1)
  7     FORMAT (1H  I3,7F16.5)
        TWOPI=6.2831853
 70     READ(5,1)M,NF,LTOTAL,NO READ, NO RHO , MOM2
C       IF NO RHO =U READS RHO,  MOM2=  2 PROGRAM COMPUTES FOR RHO
        AGAIN = MOM2'
        IF(NO READ) 71,71,69
 69     READ(5,2) (XX(I),I=1, LTOTAL)
        READ(5,2) (YY(I),I=1, LTOTAL)
 71     IF(NO RHO) 79,79,73
 73     READ(5,2) (RHO(I),I=1, LTOTAL)
 79     INDEX = 0.
        LAST = M+NF -  1
        DO 72 K=NF,LAST
         INDEX = INDEX + 1
         X(INDEX) = XX(K)
         Y(INDEX) = YY(K)
 72     CONTINUE
        TIME = 1.
 78     CONTINUE
        WRITE(6,5)
        LINE=0.
        SXA=0.0
        SXB=0.0
        SYA=0.0
        SYB=0.0
        SXHA=0.0
        SYHA=0.0
        N=M-1
        AM=M
        AN=AM-1.0
        DO 10 I=1,M
        SXA=SXA +X(I)*X(I)
        AI=I
        SXB=X(I) + SXB
        SYA= SYA+ Y(I)*Y(I)
        SYB=Y(I)+ SYB
        SXHA=X(I)*(AI-1.0)+SXHA
        SYHA=Y(I)*(AI-1.0)+SYHA
 10     CONTINUE
        SH2=AN*(2.0*AN+1.0)/6.0-AN*AN/4.0
        SX2=(SXA -(SXB*SXB)/AM)/AM
        SY2=(SYA- (SYB*SYB)/AM)/AM
        SXH=(SXHA - (SXB*AN/2.0))/AM
```

```
      SYH=(SYHA - (SYB*AN/2.0))/AM
      T= SX2 + SY2
      YB=SYB/AM
      XB=SXB/AM
      XHB= AN/2.0
      B1H= SXH/SH2
      A1H= XB - B1H*XHB
      B2H= SYH/SH2
      A2H= YB - B2H*XHB
      WRITE (6,100)
100   FORMAT (1H0,11X,3HSX2,13X,3HSY2,13X2HA1,13X,2HB1,14X,2HA2,,13X,
     1 2HB2)
      WRITE (6,3)SX2,SY2,A1H,B1H,A2H,B2H
      LINE = LINE + 3
      SU=0.0
      SV=0.0
      DO 15 K=1,M
      AK = K
      U(K) = X(K)-(A1H+B1H*(AK-1.0))
      V(K) = Y(K)-(A2H+B2H*(AK-1.0))
      SU = SU+U(K)
      SV = SV+V(K)
15    CONTINUE
      DO 18 J=1,M
      AJ = J
      RS = 0.0
      CS = 0.0
      DO 16 K=1,M
      AK = K
      Z = (TWOPI*(AJ-1.0)*(AK-1.0))/AM
      Z=AMOD(Z,TWOPI)
      SZ=SIN(Z)
      CZ=COS(Z)
      RS = RS + (U(K)*CZ + V(K)*SZ)
      CS = CS + (V(K)*CZ - U(K)*SZ)
16    CONTINUE
      R(J) = RS/AM
      C(J) = CS/AM
18    CONTINUE
      SA = 0.0
      DO 19 J=1,M
      A(J) = R(J)**2 + C(J)**2
      SA = SA + A(J)
19    CONTINUE
      Q = SA/T
      RRR = (1.-Q)*100.
      DO 20 J=1,M
      S(J) = A(J)/SA
      P(J) = 100.*Q* S(J)
20    CONTINUE
      WRITE (6,101)
101   FORMAT (1H0,3HJ/H, 8X,4HX(H),10X,4HY(H),10X,4HU(H),10X,4HV(H),10X,
     1 4HR(J),10X,4HC(J),10X,4HS(J),10X,4HP(J) )
      LINE = LINE + 2
```

```
      DO 22 J= 1,M
      I=J-1
      WRITE (6,4)I,X(J),Y(J),U(J),V(J),R(J),C(J),S(J),P(J)
      LINE = LINE + 1
      IF (LINE - 54) 22,22,21
21    WRITE (6,5)
      WRITE (6,101)
      LINE = 2
22    CONTINUE
      WRITE (6,102)SU,SV,SA,RRR
102   FORMAT (1H0,14HSUM U SUB H = ,F16.5,16X,14HSUM V SUB H = ,F16.5,
     1 16X,4HV = ,F16.5//12H R=100(1-Q)=  F16.5)
      LINE = LINE + 2
      IF (LINE +M -51) 24,24,23
23    WRITE (6,5)
      LINE = 0
24    CONTINUE
      II = 0
      JPHI=0
      DO 50 L=1,M
      H = L
      IF (JPHI- 3) 25,45,25
25    JPHI=JPHI+1
      SUH = 0.0
      SVH = 0.0
      DO 30 J=1,M
      AJ = J
      Z = TWOPI*(H-1.0)*(AJ-1.0)/AM
      Z= AMOD(Z,TWOPI)
      SZ = SIN(Z)
      CZ = COS(Z)
      SUH = SUH +R(J)*CZ - C(J)*SZ
      SVH = SVH +C(J)*CZ + R(J)*SZ
      JJ=II+J
      UH(JJ  ) = SUH
      VH(JJ  ) = SVH
30    CONTINUE
      II = II + M
      GO TO 50
45    L3=L-4
      L2=L-3
      L1=L-2
      IF (LINE + M - 51) 452,452,451
451   WRITE (6,5)
      LINE = 0
452   WRITE (6,103)L3,L2,L1
103   FORMAT (1H0,12HH SUBSCRIPTS,6X,I3,30X,I3,29X,I3/3H J ,9X,5HSU(H),
     111X,5HSV(H),11X,5HSU(H),11X,5HSV(H),11X,5HSU(H),11X,5HSV(H))
      LINE = LINE + 3
      DO 48 J=1,M
      I= J-1
      JJ=M+J
      JJJ=2*M+J
      WRITE (6,7)I,UH(J),VH(J),UH(JJ ),VH(JJ ),UH(JJJ  ),     VH(JJJ  )
```

```
      LINE = LINE + 1
      IF (LINE - 54) 48,48,47
47    WRITE (6,5)
      WRITE (6,103)L3,L2,L1
      LINE = 3
48    CONTINUE
      JPHI=0
      II  = 0
      GO TO 25
50    CONTINUE
      L3= M-2
      L2= M-1
      L1= M
      IF(JPHI)70,70,51
51    IF (LINE + M - 51) 53,53,52
52    WRITE (6,5)
53    GO TO (54,56,58),JPHI
54    WRITE (6,103)L1
      DO 55 J =1,M
      I = J - 1
      WRITE (6,7)I,UH(J),VH(J)
55    CONTINUE
       GO TO 80
56    WRITE (6,103)L2,L1
      DO 57 J =1,M
      I = J - 1
      JJ=M+J
      WRITE (6,7)I,UH(J),VH(J),UH(JJ ),VH(JJ )
57    CONTINUE
       GO TO 80
58    WRITE (6,103)L3,L2,L1
      DO 59 J=1,M
      I=J-1
      JJ=M+J
      JJJ=2*M+J
      WRITE (6,7)I,UH(J),VH(J),UH(JJ ),VH(JJ ),UH(JJJ  ),      VH(JJJ
59    CONTINUE
80    IF(AGAIN-TIME) 77,77,75
75     TIME = 2.0
       INDEX = 0
       DO 76 K=NF,LAST
        INDEX = INDEX + 1
        X(INDEX) = XX(K) *RHO(K)
76      Y(INDEX) = YY(K) *RHO(K)
      GO TO 78
77    IF(0) 60,70,60
60    CALL EXIT
      STOP 7777
      END
```

TABLE A-2

MONTGOMERY, ALA. JUNE 8, 1956 1000 HRS. (2-25 KM)

SX2	SY2	A1	B1	A2	B2
14.15880	59.38204	0.72905	-0.30411	-12.31460	0.13966

J/H	X(H)	Y(H)	U(H)	V(H)	R(J)	C(J)	S(J)	P(J)
0	0.69600	-4.95100	-0.03305	7.36360	-0.00000	-0.00000	0.00000	0.00000
1	2.50900	-6.53500	2.08406	5.63994	-0.66212	5.58368	0.46375	42.99090
2	3.52700	-4.85400	3.40617	7.18128	0.60549	-3.12177	0.14833	13.75023
3	1.65800	-1.11800	1.84128	10.77763	0.31760	-0.34275	0.00320	0.29690
4	1.04200	-5.90900	1.52939	5.84697	-0.62011	-0.55937	0.01023	0.94836
5	2.92400	-9.56300	3.71550	2.05331	0.67877	-0.32041	0.00826	0.76609
6	-3.70800	-11.41300	-2.61239	0.06365	0.16611	-0.51087	0.00423	0.39240
7	-7.45200	-15.27900	-6.05228	-3.94200	-0.46380	0.61654	0.00873	0.80939
8	-7.48800	-21.74700	-5.78417	-10.54966	-0.13281	0.60844	0.00569	0.52738
9	-7.32100	-16.44400	-5.31306	-5.38632	0.44329	0.02565	0.00289	0.26810
10	-4.91800	-18.35300	-2.60596	-7.43497	0.02259	0.10435	0.00017	0.01550
11	-3.59800	-22.71700	-0.98185	-11.93863	0.25552	0.11550	0.00115	0.10692
12	-5.19800	-24.45400	-2.27774	-13.81529	-0.11651	-0.34505	0.00195	0.18035
13	0.	-23.00000	3.22437	-12.50094	0.13251	0.23587	0.00107	0.09953
14	-1.09900	-20.97100	2.42948	-10.61160	0.11132	0.36426	0.00213	0.19727
15	-1.70600	-13.89600	2.12659	-3.67626	0.29245	-0.03007	0.00127	0.11753
16	0.47100	-8.98800	4.60770	1.09208	-0.63072	0.36113	0.00775	0.71827
17	-1.04200	-5.90900	3.39881	4.03143	0.68017	0.37661	0.00887	0.82195
18	-6.25200	-6.47400	-1.50708	3.32677	0.02112	0.83544	0.01024	0.94969
19	-5.56300	-2.24800	-0.51397	7.41311	-0.47952	0.82041	0.01325	1.22791
20	-2.97100	0.41800	2.38214	9.93946	-0.10741	-0.83048	0.01029	0.95352
21	-3.99000	-0.27900	1.66724	9.10280	-1.52722	-0.33035	0.03581	3.31999
22	-6.89400	-5.78500	-0.93265	3.45714	-0.47445	-0.13195	0.00356	0.32977
23	-10.06400	-6.53600	-3.79854	2.56649	1.45470	3.83875	0.24719	22.91538

SUM U SUB H = -0.00000 SUM V SUB H = -0.00002 V = 68.17481

R=100(1-Q)= 7.29667

APPENDIX IV-B

PREDICTION EQUATIONS

This Appendix discusses some of the procedures required for the satisfactory solution of Eq. (6.1).

$$\hat{B} = (G\,G^T)^{-1}\,G\,S^T \ ,$$

expressing the matrix $\hat{B}$ of regression coefficients as a function of two matrices, G and S. These represent, respectively, all except the last and all except the first of a series of $q + 1$ vector wind soundings through a given height interval, at a given location, in a specified month or "season".

Each sounding is represented by the coefficients of an augmented complex Fourier polynomial, computed by the FORTRAN IV program given in Appendix A. Before such computation, however, several preliminaries may be required. These may involve the separation of a long sequence of soundings, such as for several years, into appropriate "seasons," and provision for overlap of one or more soundings from season to season.

The number of soundings for each season, as well as the lowest and highest height levels for which data are to be used, must be specified. So many steps and so many variable are required for the entire computation that almost all letters of the alphabet are used, requiring in some place the use of double letters. These indicate single numbers, and not multiplication.

Schematically, the steps to be followed are:

A. Enter basic sounding and convert, if needed, to x and y.
   1. Introduce complex Fourier analysis subroutine (App. A.)

B. Compute coefficients for regression.
   1. Printout Fourier representation (optional).

C. Index Fourier coefficients.

D. Compute regression coefficients.
   1. Compute and printout standard error of estimate of estimated regression coefficients (optional).
   2. Compute and printout multiple regression coefficients of estimated regression coefficients (optional).

E. Form, store on tape, and printout regression coefficient matrix $\hat{B}$.

Some of the instructions for preparation of a program to accomplish Step A are:

1. Read basic data tape and label each sounding as T_L for L = 1, 2, ..., QQ.

2. Divide into E intervals by $ZZ = (QQ - PP)/E$, where PP is the the number of soundings to be used as overlap. Thus $L = I + MM(ZZ)$ for $MM = 0, 1, \ldots, E - 1$, and $I = 1, 2, \ldots, ZZ + PP$.

3. Reindex each of the E segments on I by $I = L - MM(ZZ)$ for each MM.

4. Call in $ZZ + PP$ soundings, converting if needed from

   v_h, θ_h to x_h, y_h by $x_h = - v_h \sin \theta_h$ and $y_h = - v_h \cos \theta_h$ and store on tape for all h.

5. Specify the atmospheric height interval, compatible with the basic input, and reindex soundings from

   $h = TT, \; TT = 1, \ldots, \; WW - 1, \; WW$ to $H = 0, 1, \ldots, N$.

6. Printout basic wind data with date and time.

Step B requires the FORTRAN IV program given in Appendix A. Only the computed trend removal and complex Fourier coefficients are used in the next step, but an option should be provided for the computation and printout of part or all of the quantities in the basic program, including the date and time of each sounding.

The complex Fourier program is applied to each of the $q + 1$ soundings to be used for estimating the regression coefficient matrix B. The linear trend and Fourier coefficients are reindexed and stored.

For Step B, the constant and linear terms are labled A1, B1 for x and A2, B2 for y. The real and imaginary parts of the Fourier coefficients, from Step B, are called R(J) and C(J), respectively, and are converted to a single sequence as g_{ki} and stored.

1. Introduce the dummy variable $g_{1i} = 1$ for all i.

2. Renumber $(A1)_i$ to g_{2i}, $(B1)_i$ to g_{3i},
   $(A2)_i$ to g_{4i}, $(B2)_i$ to g_{5i}.

3. Exclude $R(0) = C(0) = 0$ and convert, for $j = 1, 2, \ldots, N$,

   R(J) to g_{ki} for $k = 2j + 4$,

   C(J) to g_{ki} for $k = 2j + 5$.

   Consequently, R(N) is $g_{p-1,i}$ and C(N) is $g_{p,i}$.

4. Form and store the matrix G* which is $p \times (q + 1)$.

An option should be provided for reducing the number of Fourier coefficients by eliminating certain rows from G* and reindexing.

For Step D, matrices G and S, each $p \times q$, are formed by eliminating the last and first columns, respectively, from G*. Then $G G^T$ $(p \times p)$ is

computed, inverted, and the inversion stored and checked for validity by verifying that $(G\,G^T)\,(G\,G^T)^{-1} = I$, the indentity matrix. Suggested steps are:

1. Call in each row, in turn, of S, starting with the second, for $k = 2$. Identify this row vector as S_k^T. Transpose to a column vector S_k and form the $p - 1$ products, one for each k, $G\,S_k$, which are $(p \times q)\,(q \times 1) = (p \times 1)$.

2. Premultiply each product by $(G\,G^T)^{-1}$ to obtain

$$\hat{B}_k = (G\,G^T)^{-1}\,G\,S_k \quad (p \times 1)\,.$$

3. For each k from 2 to p, compute the standard error of estimate of the kth Fourier coefficient (Step D-1) and the multiple correlation coefficient (Step D-2) of the kth Fourier coefficient with all the coefficients of the preceding sounding as

$$\sigma_{Ek}^2 = \frac{S_k^T\,S_k - (G\,S_k)^T\,\hat{B}_k}{q - p}$$

$$R_k = \frac{(G\,S_k)^T\,\hat{B}_k - q\,(\bar{S}_k)^2}{S_k^T\,S_k - q\,(\bar{S}_k)^2}$$

where $\bar{S}_k$ is the mean of the q elements g_{ki} or S_k, or

$$\bar{S}_k = \frac{1}{q}\sum_{i=2}^{q+1} g_{ki} \quad \text{and} \quad S_k^T\,S_k = \sum_{i=2}^{q+1} g_{ki}^2\,.$$

4. Array the $p - 1$ vectors $\hat{B}_k$ as the columns $\hat{B}_k$ of the matrix $\hat{B}$ which is $p - 1 \times p$, with elements $\beta_{m\ell}$, where $m = k - 1 = 1, 2, \ldots, p - 1$, and $\ell = 1, 2, \ldots, p$.

This matrix $\hat{B}$ may now be used to predict G_{i+1}, the vector of Fourier coefficients of the $i + 1$st sounding (preceded by 1), from G_i, the corresponding vector for the ith sounding, by

$$G_{i+1} = \hat{B}\,G_i\,.$$

APPENDIX IV-C

PREDICTION AND VERIFICATION

Given a suitable matrix $\hat{B}$ of regression coefficients (developed as outlined in Appendix B) and a wind sounding, the general procedure proposed for predicting the subsequent wind sounding is to express the sounding by the coefficients of its representation as an augmented complex Fourier polynomial, arrange the coefficients into proper form, multiply them by the B matrix, and convert the predicted Fourier coefficients into wind components at the atmospheric levels of interest, which need not be the same as those for which the first sounding is given.

Steps in procedure are:

1. Enter the sounding (App. B, Step A)
2. Compute the trend and Fourier coefficients (App. A)
3. Reindex coefficients (App. B, Step C) as g_1, g_2, ..., g_p, to form the vector G_i.
4. Multiply $\hat{B}\,G_i$ to obtain the predicted coefficients s_m for $m = 1, 2, \ldots, p-1$.
5. Reindex the coefficients into Fourier program notation as

$$s_1 = A1, \quad s_2 = B1, \quad s_3 = A2, \quad s_4 = B2;$$

$$s_5 = R(1), \quad s_7 = R(2), \ldots, s_{p-2} = R(N);$$

$$s_6 = C(1), \quad s_8 = C(2), \ldots, s_{p-1} = C(N).$$

6. Using R(j) and C(j) compute the predicted zero-mean wind components for any height H, or for $H = 0, 1, \ldots, N$ from

$$u_H = \sum_{j=1}^{N} \left[R(j) \cos \lambda j H - C(j) \sin \lambda j H \right]$$

$$v_H = \sum_{j=1}^{N} \left[C(j) \cos \lambda j H - R(j) \sin \lambda j H \right]$$

where $\lambda = 2\pi/(N+1)$.

An optional check on the initial computations is verification that

$$\sum_{H=0}^{N} u_H = \sum_{H=0}^{N} v_H = 0$$

7. From each u_H , v_H , or only for those of interest, find

$$x_H = u_H + (A1) + (B1)\,H\,, \qquad y_H = v_H + (A2) + (B2)\,H\,.$$

8. Convert the subscript H to the actual height indicator h .

9. Verification of the prediction, after the observed values x_h' , y_h' the $i + 1$st sounding are obtained, may be in terms of any or all of the following:

   a. The scalar difference, $x_h - x_h'$, $y_h - y_h'$;

   b. The vector difference,

$$\sqrt{(x_h - x_h')^2 + (y_h - y_h')^2}\ ;$$

   c. The percent difference, $100 \left| \frac{x_h - x_h'}{x_h'} \right|$ and $100 \left| \frac{y_h - y_h'}{y_h'} \right|$.

REFERENCES

1. Hess, Seymour L., Introduction to Theoretical Meteorology, Henry Holt & Co., 362 pp., 1959.

2. Humphreys, W.J., Physics of the Air, McGraw Hill, 654 pp., 1929.

3. Lighthill, M.J., Fourier Analysis and Generalized Functions, Cambridge University Press, 79 pp., 1960.

4. Kendall, M.G., Rank Correlation Methods, Hafner Pub. Co., 199 pp., 1962.

5. "Atlantic Missile Range Reference Atmosphere for Cape Kennedy, Florida", Part I, (Prepared by the Range Reference Atmosphere Committee Meteorological Working Group of the Inter-Range Instrumetation Group), IRIG Document 104-63, 1963.

6. Hannan, E.J., Time Series Analysis , Methuen and Co. Ltd., 152 pp., 1960.

7. David, F.N., "Tables of the Ordinates and Probability Integral of the Distribution of the Correlation Coefficient in Small Samples", Issued by the Biometrika Office, University College, London. Cambridge, The University Press, 1938, 55 pp.