[illegible] RESEARCH AND DEVELOPMENT PAPER

[illegible]

A PRINCIPAL COMPONENT ANALYSIS OF RECORDS AT A SINGLE CLIMATOLOGICAL STATION

E. J. White and D. K. Lindley

R. & D. 41/73

CONTENTS

TABLES

FIGURES

I Introduction

1. Background

Research currently undertaken by the Woodland Section of the Merlewood Research Station of the Nature Conservancy into British semi-natural woodlands has been developed from an association analysis of 2463 sites (Bunce, 1970) carried out on presence or absence of plant species. The classification of these sites has been conveniently referred to 103 groups, and more intensive recording of site variables on a representative wood from each of the 103 categories is being carried out.

While it is reasonably possible to sample vegetative and soil characters during the course of one or two visits, sampling of climatic variables presents a greater problem. Basically, five different difficulties occur:

a. Woodland is a deep boundary layer between the atmosphere and the soil and so presents difficulties in the vertical placing of the instruments especially compared with the siting of the standard instruments of the Meteorological Office at a height of 1.3 m over mown turf. Standardisation with data collection outside this study is desirable in case comparisons are required.

b. British woodland is mostly in small parcels which generate local micro-climates.

c. Even with continuous woodland greater friction and hence turbulence occurs than with mown turf.

d. Even supposing these problems could have been overcome and instruments had been set up gathering data would have been expensive in personnel time and equipment.

e. Data collection in this way would not have projected backwards in time and it is past weather which has partly determined the present vegetation rather than present weather especially for trees. It is hoped that the problem will be resolved by an analysis of United Kingdom data from standard meteorological stations, spaced as evenly as possible, over as long a time as possible and using trend surface plotting to predict values for intermediate locations. Principal component analysis will be used to reduce the data and the final estimates will be for values of principal components and/or variables. It will be possible to test predictions from the trend surface model against values at meteorological stations other than the ones used in this analysis. The predictions can later be used as regressor variables in multiple regressions with a plant production parameter as the dependent variable.

There are several approaches to the problem of examining plant/animal response to climate. First, response can be compared with one variable (Wareing 1956) or with a few selected meteorological variables by the calculation of simple correlation coefficients (Hiley and Cunliffe, 1923, Millar, 1964a). If carried to its logical conclusion this procedure results in a large array of regressor variables with perhaps some of the day or week before and the addition of squared or cubed values (Fritts 1960).

Such arrays are then frequentl, subjected to multiple regression analysis and suffer from the attendant disadvantages arising from intercorrelations between the regressor variables and lack of homogeneity in the distribution in the values (Jeffers pers. comm.). An alternative approach has been to calculate an index of climate of which estimates of evapotranspiration have been the most common and to compare response with this index. The approaches of carrying out separate principal component analyses of tree growth variables and of site variables in several groups if necessary followed by calculating correlation coefficients between the components of the separate analyses (Fourt etal, 1971), or of carrying out a principal component analysis of the meteorological variables followed by an orthogonalized regression against plant response (White, 1970) are seen as alternatives which do not suffer from some of the disadvantages of the other methods. At the same time, it has the advantage that large numbers of variables can be screened and reduced objectively. Also, the proportion of variability accounted for can be accurately determined.

Interest in meteorological variables as regressor variables in regressions of plant response on climate does not necessitate further understanding of the variables or of particular responses, but has aided their interpretation as well as helping to place this study into context with others.

It has been pointed out (Seal, 1964) that a variance-covariance matrix should not be used in an analysis of variables measured in different units, as, if the units are changed or variables are standardised, different results are obtained. If actual values of several variables are used, or departures from means, or standardised departures, Kutzbach (1967) states that a cross-product matrix a covariance matrix or a correlation matrix is obtained respectively, and that the principal components will most resemble the observed field the departure fields, or the standardised departure fields respectively. The use of standardised variables determines the extent to which each variable at each station is of importance in determining the form of representation.

In this analysis the matrix of correlation coefficients is obtained from the variance-covariance matrix by dividing the corrected sums of products by the standard deviations corresponding to the appropriate two variables the results being equivalent to using standardised variables.

Thus several types of principal component analysis have been used in meteorological studies. Principal component analysis of fields of one variable have been carried out by Craddock and Flood (1969), Lorenz (1956), Lund (1963), Perry (1970), Sellers (1968) and Stidd (1967).

Shorr (1956) and Veitch (1970) have separated fields of each variable before carrying out a principal component analysis on each.

Christensen and Bryson (1966), Steiner (1965), Kutzbach (1967), Fourt et al (1971) Newnham (1968) and White (1970) have analysed fields of several variables using matrices of correlation coefficients. Christensen and Bryson examined variation with time (during January) at two stations, while White examined variation at one station, and these analyses show greater similarity in data, method and object to the present one. Steiner and Kutzbach examined geographical variation, whereas Fourt et al and Newnham examined variation in time and in space. Comparisons will be made between the results from this study and from others at the discussion stage.

2. Reasons for the Analysis of Moor House Data

It became evident that a number of advantages would result from an initial analysis of data from one station only. This paper describes such an analysis, based on data from the climatological station at Moor House an experimental site on a National Nature Reserve administered by the Nature Conservancy.

This analysis was intended to serve five purposes:

a. First, and mainly analyses were made of daily values of the variables and of five-daily monthly and quarterly means with the object of selecting the length of the intervals between sets for the complete U.K. analysis in the hope of achieving large economies in computing time and storage on magnetic tapes for that analysis Whilst it is desirable to reduce the number of variables for an given woodland site to the minimum it is necessary to have some idea how much information is lost in the process and to reach a balance between loss of information and gain in economy. The data can be reduced by using longer intervals and by taking the values of fewer principal components. It is therefore necessary to consider the nature of information lost in reducing the data. Can the information contained in data with short intervals and which is lost from data with long intervals, be regarded as background noise, and has it no greater significance?

b. The stability of the components can be considered, both seasonally and from year to year. If the principal components are stable, from month to month and from year to year, i.e. if the weightings of the variables are fairly consistent, the question arises as to whether, if the principal components are empirical functions, they have any meaning in climatological terms? For instance, do they have any meaning in terms of the different weather types identified by Lamb (1965), or do they have any meaning in terms of the phenomena which he (1971, 1972) considers determine change in our climate, namely changes in heating and cooling of the atmosphere, albedo (due mainly to cloud, snow and ice), surface temperature of the oceans, insolation to dust and pollution in the atmosphere, and output of the sun?

The meteorologist is largely concerned with atmospheric dynamics of which the variables we have chosen as significant to plants are merely part of the expression at ground level of changes in the atmosphere. The empirical orthogonal functions derived in this study are not thus thought to be direct expressions of the phenomena Lamb considers as determining climatic change. Should principal component analysis of these phenomena be attempted a canonical correlation analysis between the set of empirical orthogonal functions obtained and those obtained here would throw some light on this problem.

For the present study a comparison with Lamb (1971 1972) is a side-issue. However, it would appear to be reasonable that within the terms of this study the greater the stability of the components the more meaningful they must be.

c. The analysis provides familiarisation with the data.

d. Basic computer programs for manipulation and calculation can be developed.

e. There will be immediate 'spin-off' to other studies at Moor House notably moorland studies within the International Biological Programme, and to tree growth experiments at high elevations.

3. Moor House and its Climate

Moor House Field Station is situated in a National Nature Reserve, at 558 m (1840 feet) O.D. at latitude 54°41'N and 2°23'W in the northern Pennines. The highest ridge of the Pennines is about two miles to the west, with Great Dun Fell as its highest point at 655 m O.D. Moor House is in a gently sloping and undulating area on the headwaters of the River Tees. The vegetation is of Calluna and Eriophorum on peat with small areas of rough grass on mineral soils. The area is common land and is grazed by sheep in the summer.

The climate of Moor House and the northern Pennines has been described by Manley (1936, 1943, 1952). He found it windy, chilly and damp, corresponding with sea level in southern Iceland. While it is much as he expected, he noted that on rare occasions great extremes of weather were possible. Of interest in the field of biology, he stated that "relatively small changes in the wetness or dryness of particular months made a big difference to the profitability or otherwise of farming in such a marginal area". He reported down-slope Katabatic flow into the Eden Valley at sunset, and expected a similar flow to occur into the Upper Tees valley (now the Cow Green Reservoir site). This seems unlikely to occur frequently because of the generally high average windspeed. In 1945 Manley reported on the Helm wind of Crossfell, a down-slope wind of high velocity which occurs on the west side of the summit ridge when the windspeed at Moor House reaches Beaufort Force 3 and a direction of E to NNE. Green (1964) plotted a map of annual potential water deficit which included data from the lysimeter at Moor House, and in 1953 commented on the occurrence of a remarkably low relative humidity. Millar (1964b) described the climate in relation to the upper forest limit which he considered to be at about 564 m O.D. Pears (1968) considered the present tree line in the Cairngorms to be between 610 and 686 m whereas Pearsall and Pennington (1947) found the tree line in the Lake District between 183 and 457 m, and these low levels were thought to be due to wind exposure. Pears observed the present tree line in the South Tyne Valley at 533 m which is close to Millar's estimate. He drew attention to Simmons' (1962) observations of a tree line on Dartmoor below 396 m and considered this very low level reflected greater oceanicity (and wind exposure) and/or influence by man from Neolithic times onwards. This positive correlation between the tree line and latitude is of some interest and it is hoped that the main analysis may throw some light on it.

II The Data

Moor House data used here were held on Climatological Land Surface Punched Cards (Form 1996). Moor House being a Series-3208 Climatological Station, recording basically the same standard data (Anon (1968)) as Health Resort and Agro-meteorological Stations. The data were analysed for the ten-year period 1960-1969 inclusive.

Since the choice of data for the Moor House analysis depended directly on the choice made in the U.K. study, the reasons considered there will be briefly described: data were required for simplicity to be uniform and on one type of punched card. Ideally the data should cover the life-span of the trees for which response will be examined. The longer the period to be included in the analysis, the fewer the meteorological stations with available records, and a balance had to be found between length of time and number and eventual distribution of stations. The balance arrived at utilizes data from 73 stations from 1960-69 inclusive. There are conspicuous gaps in the uplands and it is likely that only the adoption of automatic weather stations will improve coverage in these areas. These gaps, and the cooling tendency during the decade, making the data doubtfully representative for, say, the last 100-150 years affecting present-day mature trees, are recognised as possible sources of error in regression relationships of plant response. Other changes may have occurred which affected this decade. For example, the Clean Air Act of 1956 resulted in a steady reduction of smoke and SO_2 in the atmosphere over Britain (Anon, 1971), an effect which is reflected in improved solar radiation. Plants may, therefore, have responded to the reduced pollution, as well as to increased solar radiation. In spite of these difficulties, it is felt that the approach used here is the best currently available until a longer run of data from more stations is available on the same cards, making a wider choice of data possible.

1. The Variables

The 33 variables given on the daily punched cards are listed in Table 1. 3653 cards were used in the Moor House analysis. Of the 33 variables for which columns were present on the cards, 8 were absent at Moor House and at many other stations. Run of wind from anemometer records on the cards, or from anemograph records separately, is available at only a few stations (including Moor House) but this was disregarded at Moor House for consistency with the main analysis. Of the variables remaining for the main analysis, the factor which most limited the selection of stations was the presence of a sunshine recorder.

Of the remaining 25 variables on the cards, the first six were 'labellers' which were discarded as not being meteorological variables. Present Weather and State of Ground were discarded as their values are coded, and the coded values follow several parallel lines. For example, in Present Weather, increasing heaviness of rain has codings in the 60's, snow in the 70's. Is snow "worse" than rain or are they of parallel severity? In contrast, visibility is coded, but was retained since it has a single, though logarithmic, scale.

As estimate of potential evapotranspiration by the Penman equation was added giving a total of 18 variables and 65,754 items of data. It may seem inconsistent to add a derived variable to the others, but this was done because many authors have used evapotranspiration as an index of climate to compare with plant growth. For example, Rosenzweig (1968) has successfully demonstrated a logarithmic relationship between the above-ground net annual primary production and actual evapotranspiration for a great range of sites from Arctic tundra to tropical forest.

Wind direction is coded in terms of degrees from 10° to 360° and is unusual in that values can in effect pass from the minimum to the maximum by a small change when the wind is backing from 10 to 360°, or vice versa when it is veering. Values for wind direction were coded by standard Meteorological Office procedure from estimates made on the basis of the 16-point compass, NNE becoming 20° and NE, 50°, so that the intermediate values 30 to 40° are missing, and so on round the compass rose, so that the resulting scale is discontinuous. A further complication is that "0" represents a calm, which is not a direction. This value has been retained in the analysis, but the possibility of representing wind direction in other ways is discussed under transformations of variables.

Frequency distributions for daily values of variables during 1961 are given in Figure 4. The dip in the frequency distribution for windspeed is an artifact due to the construction of the classes in the distribution: Beaufort Forces 3 and 4 with mean windspeed of 9 and 13 knots fall into frequency class 3, whereas Force 5 falls into class 5, leaving no occurrence in class 4. The dip can probably be ignored in interpretation.

Rainfall is confused with fresh snow depth, since freshly fallen snow in the rain gauge is melted and measured as rainfall.

In the case of visibility, "x" and "y" on the cards were interpreted as -2 and -1 respectively by the card-reading programs.

Frequency distributions in Figure 4 indicate that relative humidity has a negatively skewed distribution, whereas rainfall, duration of sunshine, total and fresh snow depth are positively skewed.

Other expressions derived from the variables used here have commonly been calculated, often in association with threshold values. The aim has been to find an index which shows a high correlation with plant or animal response. Examples are length of frost free period (days), length of growing season (when the mean temperature is above 42°F), degree days above 42°F, number of days of gale, and the Penman estimate of potential evapotranspiration.

As mentioned above, the last of these indices has been included in the analysis.

The other expressions could have been computed and added to the existing array. Increasing the number of variables in this way would have caused practical difficulties in computer storage requirements and in interpretation of the principal component analysis. More seriously, there are other objections to the use of such derived variables. First (and practically) each expression associated with a threshold value must be correlated with the variable it is derived from in a regression equation, with a constant depending on the value of the threshold. Hocker (1956) for instance noted a high correlation between length of the frost free period and winter temperature. When the relationship between meteorological variables or their empirical orthogonal functions with plant or animal response is examined by multiple regression, a single combined regression constant results which achieves the same end in a more economical manner.

Second, and theoretically, plant and animal response are related to the environment through an essentially multivariate situation. It is not, therefore, surprising if the calculated derived variable is not highly correlated with response. When tree growth near the Arctic Circle is limited by summer temperature, and near the Sahara by moisture, it can only be a multivariate approach that holds any hope of successfully explaining both situations, and those in between, by a single model.

2. Error Detection and Correction

Figure 1 shows the basic method adopted for the correction of errors in large data matrices, "large" being defined for this purpose as a size above which it is uneconomic to examine each item by eye. The method uses three computer programs applied to one year's daily data at a time. The program SVN4 identified items more than 4 times the standard deviation from the mean, and functions well except for variables with skewed distributions such as rainfall and snow depth, ten or so items above the mean being identified each year. These are considered in relation to experience at the particular station, and are mostly accepted. SEDA is used for identifying the set number where an erroneous maximum or minimum values occurs, for instance a negative value for rainfall.

The Meteorological Office, and Mr. A. Millar of the Nature Conservancy, in the case of Moor House data, were consulted where values appeared unacceptable, and replacements were made using the program REDA which reads the data to the particular set number. The program MMMS has been used (see Figure 2) to check that the correct variables have been selected from the basic magnetic tapes. It is recognised that some errors probably still exist in the data, and that error correction as tackled removes only those furthest from the mean. The aim can only be to keep errors down to an acceptable level in a matrix of the size involved here (with nearly 66,000 items of data). A number of card-punching errors were found but the most common error at the observer level was the appearance of odd values of relative humidity, suggesting the wet bulb had been mis-read by 5 or 10 degrees.

Blanks on cards were read as zeros by the program CTOP, and occurred for vapour pressure, relative humidity and dew point during December 1962. The zeros did not appear as exceeding 4 x SD from the mean in SVN4, but as erroneous minima.

3. Preparation

Figure 2 shows a schematic network diagram for the preparation of the final 18 variables which incorporates Figure 1. The preparation stages consisted of translation of the data into a form the computer could use, error correction, selection of variables, and the addition of the Penman estimate of evapotranspirati for which the method of computation is that given by Smith (1967). The variables from which the Penman estimate is derived are windspeed, vapour pressure, mean daily maximum and minimum temperatures and duration of bright sunshine, together with half the calculated daylength and potential solar radiation per unit area at the top of the atmosphere. The method used to calculate half the daylength is given by Frank and Lee (1966), and that for solar radiation by Sellers (1965). Methods are described further in the specifications for computer programs, PENM, SOLM and SLRM respectively. Frequency distributions of the variables were also examined since variables with two modes are best avoided in principal component analaysis.

III Method of Analysis

The scheme of analysing the data using the method of principal components is shown in Figure 3. All the work was carried out on a Digital Equipment Corporation PDP-8/I computer. The programs are all available at Merlewood for the PDP-8 and are in 4K Fortran-D language except for SLCT and CORR and EIG for up to 20 variables, which are in 8K Fortran, and CTOP which is in the machine code PAL-D.

The correlation matrix program CORR obtains the correlation matrix from the variance-covariance matrix by dividing through by the standard deviations corresponding to the appropriate two variables equivalent to the calculation of standardised variables. The program CORR functioned satisfactorily for reasonable lengths of paper tape. When tapes became very long, for instance a daily tape for 10 years would take 8 hours to read through the high speed reader, the program was modified to take one year's data at a time and to produce an intermediate output. These were combined for 10 years by the programs CORM1, CORM2 and CORF. These did not produce overall means, but this was not essential, since these were available for 5-daily data from the CORR program.

Another modification of CORR was made in the program CORA, the program used on quarterly data. Since the program CORR would not run if a variable had only zero values, CORA modified zero values of total and fresh snow by replacing with the insignificant values +0.0001 and -0.0001 on alternate dates. This modification was required particularly in the third quarter when no snow fell, but was carried out on all quarterly analyses for the sake of consistency.

IV Results and Discusion

1. Correlation matrices

An 18 x 18 half matrix of correlation coefficients, without the principal diagonal, contains 153 coefficients. In the analysis described by this paper, 63 such matrices have been calculated and presentation of all these matrices is scarcely feasible. Only a few of the more important matrices, selected to show the effect of analysing means of 5-day, monthly and quarterly data by contrast with daily data, and the stability of the results from one year to another, have therefore been presented.

The analyses made are listed in Table 3. The half-matrix of correlation coefficients derived from daily values of the ten-year period (analysis A) is presented in Table 5, and illustrates the high degree of intercorrelation present between the temperature variables.

The correlation coefficients of Table 5 were compared with those derived from the other analyses in 53 scatter diagrams. Examples are given in Figure 5, with significance at the 0.05 level of probability indicated.

a. Stability of correlation matrices derived from data representing different time intervals.

A method was required to examine the degree of stability of the correlation coefficient between any given pair of variables within a series of matrices of correlation coefficients. The series examined were the ten annual matrices derived from daily data, the ten matrices derived from daily data for each quarter of each year, and the matrices derived from 10 years daily data, and 5-daily, monthly and quarterly means. The method used consisted of examining the range of values for the correlation coefficient for each pair of variables by means of scatter diagrams, some examples of which are shown in Figure 5, and allocating each coefficient to a stability class, as follows:

Class	Description of Each Correlation Coefficient over a Series of Matrices
Stable	All positively significant, all non-significant, or all negatively significant.
Partly stable	Varying from positively significant to non-significant or non-significant to negatively significant.
Unstable	Varying from positively significant to negatively significant.

Significance was only examined at the single 0.05 level of probability. The choice of a single criterion, and of this particular criterion, may be open to question. The number of correlation coefficients in the Stable Class in any one series was then used as a stability index for that series. Clearly, to have examined the degree of change over several levels of probability would have presented problems in choice of weightings, and in subsequent reduction to one value.

The method was used for the comparison of the stability of the series of matrices listed above, with a possible maximum value of the stability index of 153 for the 18 x 18 matrices without the principal diagonal, i.e. without duplicates. The method was also used to compare the stability of each variable, when examining 17 correlation coefficients for each of the 18 variables, i.e. including duplicates, giving a possible maximum value for the stability index of 306. The indices for matrices and variables are not therefore comparable.

The stability of the matrices of correlation coefficients is used as an index of the reliability of the analyses. Stability can also be examined from the eigenvectors. Comparisons of the two methods are discussed later.

Values of stability indices are shown in Table 6. The similarity between half-matrices of correlation coefficients for 10-year runs of daily data, and of 5-day, monthly and quarterly means (analyses A, B, C and D) is shown by a relatively high index of 114. The high stability suggests that any one of these sets of data could be chosen for analysis without altering the meaning of the resulting analysis very much, as judged by this criterion.

The fact that seasonal differences are greater than annual differences is expressed by the fairly high value (98) of the stability index for the 10 matrices of correlation coefficients for the series of annual analyses of daily data (analyses E to N). This expresses in other words the wide range of values and the closeness of annual means.

The range of most variables is reduced in each quarter. This appears to be the reason for the unexpectedly low values for the stability indices for the 10 matrices of correlation coefficients for the series of analyses of daily data for any one quarter of each year (analyses O, P, Q and R).

Some correlation coefficients change from one season to another. This is expressed in the low value of the stability index of 66 for the series of 4 matrices of correlation coefficients resulting from 10 yearly analyses (S, T, U and V) for daily data of each quarter. For example, cloudiness at 0900 hours and dry bulb temperature are positively correlated in the first and last quarters, and negatively correlated in the second and third quarters. In other words, anticyclones with clear skies are associated with cold weather in winter, and with warm weather in summer. Similar changes occur, between seasons, for other variables which are highly correlated with cloudiness and dry bulb temperature. Most yearly correlations between cloudiness and dry bulb temperature are insignificant (in analyses E to N), being calculated for periods which are too long to be sensitive to seasonal variation. In the 10-yearly analyses of daily, 5-daily, monthly and quarterly data (analyses A, B, C and D) correlations between cloudiness and dry bulb temperature are all insignificant, and thus fall into the stable class.

b. Stability of variables

Values of the stability indices for the variables are given in Table 7. Wind direction and speed are least stable, followed by rainfall. This is in spite of the fact that the coefficients of variation for wind direction and speed are similar to those for most variables (see Table 2), though the difficulties associated with the quantification of wind direction have already been mentioned. Wind direction and speed are both liable to sampling error, in that both are very variable over short periods of time, and both are assessed more or less instantaneously. The low stability of rainfall is expected since it is positively skewed, and its coefficient of variation is high (94.8%).

2. Interpretation of Principal Components Derived from Data Representing Different Time Intervals

a. Eigenvalues and the percentage variation accounted for in 10-yearly analyses.

The eigenvalues, percentage variation accounted for, and cumulative percentages are shown in Table 8 for the first 5 principal components of 10-yearly analyses (A, B, C and D) of daily data and of 5-daily, monthly and quarterly means. The percentage accounted for and cumulative percentages are graphed in Figure 7.

The curves are consistent, and it is evident that, the longer the period represented by each data set on which the analyses are based, the greater the percentage of variation accounted for by the first principal component, later components showing slight reductions. In other words, the seasonal temperature effect expressed by the first component becomes more dominant. The calculation of means has reduced the variation present in the data in proportion to the length of the period for which they have been calculated. The second graph in Figure 7 shows that the longer the period represented by the means, the greater the percentage variation accounted for by a given number of principal components, and the fewer the components required to explain a given percentage of variation. That part of the variation which has been removed by the calculation of means will be discussed later.

b. Eigenvectors in 10-year analyses

Eigenvectors for 10-year analyses (A, B, C and D) of daily data and of 5-daily, monthly and quarterly means are given in Table 8. Values greater than +0.7 or less than -0.7 are underlined and indicate variables with the most important weightings (Jeffers 1967).

The first principal component for daily data represents temperature with heavy weightings of all temperature variables, and of vapour pressure and dew point, but not of relative humidity percent. The temperature variables themselves express the seasonal effects, and it has already been noted that these show high intercorrelations in the matrix of correlation coefficients. Calculated P.E. does not appear important.

As the length of periods represented by the data sets increases, additional variables show important weightings in the first principal component. These are visibility, hours of sunshine, snow depth and calculated P.E., in other words the first component becomes a more general expression of energy. These changes are shown in Figure 8.

The second component expresses atmospheric moisture, the important variables being cloud and relative humidity contrasted with visibility and sunshine, in the case of daily data and 5-daily means. Good visibility at Moor House is associated with dry weather, which may not be the case in the lowlands. When the data sets consist of means calculated from longer periods, the second component changes in character, and represents cloudiness and rainfall only, and the other variables which had been important now appear with increased weightings in the first component. As each component is calculated on the residual matrix after calculation of the previous one, a change in the character of the first component necessarily causes changes in the subsequent ones.

The third principal component for daily data is an index of wind direction and speed, i.e. high windspeed is associated with westerly winds. This was expected since these variables showed a low, but significant, positive correlation coefficient in the matrix of correlation coefficients (Table 5). When the data sets consist of 5-daily and monthly means, rainfall is included as an important variable. But when they consist of quarterly means, the third component changes in character considerably, becoming an index of snowfall, rainfall having already been included in the second component.

There is one exceptional case where the same variable shows important weightings in two successive components calculated from the same data matrix. This is the case of rainfall in the second and third components calculated from monthly means.

The fourth component expresses snowfall, except when calculated from quarterly means, in which case it e presses windspeed.

The eigenvalue of the fifth component is under 1, and is therefore assumed to account for an insignificant part of the variation. In the case of daily data, rainfall surprisingly does not appear with an important weighting in any component before this one. The fifth component for variables consisting of means expresses windspeed contrasted with wind direction i.e. high windspeed associated with north-easterly winds, cloudiness having a high weighting in the case of the quarterly means.

The relative stability between matrices of correlation coefficients calculated from daily data and 5-daily, monthly and quarterly means has already been noted. The components show some change, but are broadly explainable in the same terms. The first represents energy, the second dampness, the third windiness (but snowfall for quarterly data), and the fourth snowfall (but windspeed for quarterly data).

c. Value of the components.

The quarterly values of the empirical orthogonal functions are listed in Table 11 and are shown graphically in Figure 9, and monthly values are shown in Figure 10. The values of the first component show a strong seasonal trend. The cold winters 1962/3 and 1968/9 correspond with the lowest points. Other comparisons with observed weather pattern will be discussed in a later paper describing a cluster analysis of the results.

d. Choice of length of period represented by each data set.

Two criteria govern the choice of length of period represented by each data set for the U.K. analysis:

1. Consistency of the components, which depends on conclusions from this initial study. Since the components are broadly explainable in the same terms, whatever length of time each data set represents, the choice seems quite open.

2. The number of values of orthogonal functions required for each site, where the dependence of vegetation response is to be examined. These values will be used as regressor variables, with values of principal components of soil variables, local topography, and possibly pollution. Therefore the fewer the better, and quarterly data seem preferable, judged by this criterion.

e. Division of the year into quarters.

The mean monthly values of the principal components are shown in Figure 10. In calculating quarterly means in this study, each year has simply been divided into four. It is evident from the monthly values of the first principal component in Figure 11, which is the main expression of the seasonal affect, that calculating quarterly means based on a first one calculated from December, January and February, would maximise the variability more fully than a calculation based on a first one from January, February and March. This would give a division of the year into quarters based on multivariate data. Bryson and Lahey (1958) worked mainly on North American data, and divided the year into divisions where the greatest differences in weather appeared to occur. Their seasons were winter (1 November to 21 March), spring (21 March to 25 June), summer (25 June to 21 August), and autumn (21 August to 1 November). The late Mr. N. C. Helliwell of the Meteorological Office has pointed out that each season could be said to commence at different times, in terms of different individual variables. For example spring starts when earth temperature at 1 foot has been above 42°F for 5 days; the water year starts with autumn rains when the soil water deficit is removed. We are concerned with plant response to climate, and division of the year into quarters for this purpose should ideally be based on biological criteria: that division should be used which accounts for the maximum variation in response A of species x. Since this approach is likely to be demanding in computer time, and since each response of each species is likely to be different, and since measures of such dependent variables have not yet been carried out, the practical approach would appear to be to divide the year into four equal quarters, with a first quarter comprising December, January and February.

f. Principal components resulting from annual analyses of daily data

Considerable consistency, suggesting stability, occurs in the standard eigenvectors of variables which are important, and even for those which are not, in the first component, in annual analyses of daily data (analyses E to N), which are shown in Figure 8. A similar picture emerges for the second component. In the third one, however, large oscillations occur in variables 2 and 3, with respect to variables 16 and 17, suggesting instability. Increasing instability occurred in the fourth component, due to the same two pairs of variables again, but now involving the other variables, with more instability again in the fifth component. Several attempts were made to remove the instability by trying various transformations of the variables, discussed later. Since the first two components, which are the stable ones, account for over 60% of the variation, the picture is generally one of reasonable stability.

8. Principal components resulting from 4 analyses of cumulated daily data of each quarter for the 10-years (analyses S, T, U, V).

Apparent instability between matrices of correlation coefficients calculated from daily data for each quarter, cumulated for the 10 years, has already been noted, and has been commented on in the instance of the correlation coefficients between cloud amount and dry bulb temperature, and the variables associated with these.

The question arises whether this relative instability appears in the eigenvectors listed in Table 9. The first components of the four quarterly analyses are again largely temperature, the values being shown in Figure 8. When the eigenvector for a variable changes from an 'unimportant' values (less than 0.7 from the mean) to an 'important' one, its actual change in value may be quite small, suggesting that taking a single value as the criterion of importance is a necessary simplification but nevertheless an oversimplification.

In the case of eigenvectors of the first component derived from annual analyses of daily data (see Figure 8), 80% of variables showing important weightings in at least one analysis (i.e. those giving meaning to the components), show importance throughout the series, i.e. in this respect they are stable. In the case of eigenvectors of the first component derived from 10-year analyses of daily data, and of 5-daily monthly and quarterly means, 64% of those showing importance in at least one analysis are stable.

Therefore the relative instability noticed between the matrices of correlation coefficients calculated from daily data of each quarter cumulated over 10 years can be identified again in the eigenvectors of the first component, using a single criterion to determine whether the values are 'important' or 'unimportant'. Figure 8 also indicates that this is accompanied by changes over a greater range of values in the eigenvectors generally. These differences are reflected in the values of the eigenvectors of the higher numbered components.

The first component calculated from daily data for each quarter cumulated for 10 years again expresses temperature, with minor variations. Earth temperature has high weightings in the second and fourth quarters when it varied most. Grass minimum temperature does not vary much in the second quarter and is unimportant. Dry bulb and maximum temperature vary little in the third quarter and are unimportant.

The second component is explained by dampness, cloud and relative humidity, being contrasted with visibility, sunshine and P.E. The last variable is no longer important in the first component, no doubt due to its reduced range in each quarter. Dry bulb, maximum and grass temperatures have important weightings in those quarters which correspond with the quarters where they are unimportant in the first component. Earth temperature in the third quarter however is not important in any of the first 5 components.

The third and fourth components are again mainly explained by unstable contrasts between the pairs of variables, wind direction and speed, and fresh and total snow depth, the second pair being absent in the third quarter.

Te fifth component is near the significance level, and is explained by different variables for each quarter for which it was calculated. Background noise is the suggested explanation.

3. Transformations of Variables

Apparent instability of the third principal component in many of the analyses led to an examination of the structure of the variables concerned (i.e. daily values for 1961), especially wind direction, which has already been discussed as being exceptional in the section on variables, although its coefficient of variation in the case of 5-daily means for 1960-69, was quite acceptable (0.296). Removing the 13 calms in the 1961 daily values reduced the coefficient of variation from 0.445 to 0.394 (see Table 4), reduction in the coefficient of variation being regarded as an indication of behaviour rather than as an end in itself. Two methods seemed possible for overcoming the objection that a small swing could be associated with a change from maximum to minimum and vice-versa. The first consisted of using the cosine transformation.

The second consisted of replacing wind direction (degrees) by two new variables 0-180 degrees north and 0-180 degrees east. The correlation coefficient between consecutive sets remained fairly constant. Both transformations, especially the cosine, show increases in the coefficient of variation over 0.394, and so the original variable was retained for this analysis.

In the case of the analysis of data for the whole of the U.K. and the examination of the dependence of plant response, the use of the two variables 0-180 degrees north and 0-180 degrees east for wind direction may have some advantages. These new variables may be the best expressions to associate with dependence of plant response on varying topography, different aspects being expressed by the same two variables. From the point of view of pure meteorology, better expressions might be 0-180^{o}NE, and 0-180^{o}NW, the first possibly maximising variation from tropical or polar air streams, and the latter possibly doing the same for continental to maritime air streams. The effect of producing two variables expressing wind direction on matrices of correlation coefficients and principal components was not examined.

The effect of the transformation $\log_e$ (variable + 10.0) was examined on daily values for 1961 of the skewed variables relative humidity, rainfall, hours of sunshine and total and fresh snow depth. This reduced the coefficient of variation of each, but the same distributions occurred within the reduced range. The effect on eigenvalues and eigenvectors was insignificant. Since the calculation of matrices of correlation coefficients is in effect based on standardised variables, this was to be expected.

It seems logical that variables should be chosen for their meaningfulness for the purpose in hand (investigating the causes of variation in plant response) and that any resulting instability of the third and fourth components should be tolerated as not being as serious as instability in the first and second components.

4. Selection of Variables

There are several criteria for the choice of variables. Availability and ease of measurement have usually assumed importance, and are the reasons for the fact that a third of the variables present are expressions of temperature. These reasons do not optimize the choice which should be determined by the purpose of the analysis, though some compromise is often necessary. Whether the variables are important to plant or animal respone and whether they contribute significantly to the overall variation in meteorological variables are the reasons which should govern the choice in this case.

It seems likely that visibility and fresh snow depth (since total snow depth is present) are unimportant to plants and animals and should be rejected, and that calculated daylength should be included, as being of fundamental importance in expressing the seasonal movements of the sun.

A separate principal component analysis was carried out on the six temperature variables, and it was evident that they could be adequately represented by wet bulb (or dry bulb) temperature, this having the greatest weighting in the only significant component (see Table 12a).

A separate principal component analysis was also carried out on the three variables representing airborne moisture at screen height, of which vapour pressure and dew point show high correlation coefficients with temperature, the results of which are given in Table 12b. These had equally high weightings in the first component, and relative humidity had a high weighting in the second component. From these considerations it seemed that the first two should be rejected. If, however, cloud amounts was added to these three variables, it did not appear as the most important variable until the third component, which was the first insignificant one.

Jolliffe (1972) discusses methods for discarding variables in a principal component analysis by removing those which contribute only an insignificant part of the whole variation present. There was little to choose between the 8 methods he describes except length of computing time. His "B4" method could conveniently be applied to this study, which retains that variable having the greatest weighting in each significant component (down to an eigenvalue of 0.71), the calculations for which had virtually all been carried out already. The results were listed in Table 10 in the case of the four 10-yearly analyses (A, B, C and D) carried out, and for the daily data for 1961, with the subjectively-based selection of variables which has been discussed above.

In all 5 cases there is a substantial measure of agreement. Only one temperature variable is retained in each analysis, cloud being the only remaining expression of airborne moisture. Windspeed and direction are retained, but only windspeed for quarterly data, together with snow depth, and rainfall in some cases.

It is encouraging that the variables retained were more or less those which Lamb (1972) considered are associated with seasonal changes in climate. These were roughly changes in surface heat with changes due to winds snow cover cloud amount, and transparency of the atmosphere. Visibility is the exception in being rejected from the 10-yearly analyses in Table 8. This may be because visibility may behave exceptionally at Moor House, which has already been suggested, and visibility at ground level is hardly a measure at Moor House of what Lamb appears to have mean by atmospheric transparency, i.e. mainly volcanic dust and pollution veils, though mist and sea salt also contribute.

It is unexpected that daylength has been rejected because as an expression of the sun's annual movement, it was felt to be fundamental. The conclusions of the separate principal component analysis (Y) of temperature variables, i.e. that these are adequately represented by one variable is confirmed by using Jolliffe's "B4" method of selection. The conclusions of analyses (Z1 and Z2) of moisture variables were not confirmed when these were included with all the other variables, those representing moisture at screen height have been absorbed in the first or second components cloud amount appearing with greater importance as the main variable of the second component (except in the case of the analysis of quarterly means). Rainfall does not appear an important source of variation in the analysis of 5-daily and monthly means

5. Comparison of Interpretations of Principal Components with Other Workers

Comparisons between interpretations of principal components derived in this and other studies is not simple. It is complicated by the choice of meteorological variables, their periodicity, and by the addition of derived variables as well as by the possible addition of topographical variables and geographical co-ordinates. Table 13 summarises interpretations from relevant studies and underlines the differences due to the above causes, though temperature (or more widely energy) appears constantly in the first component.

V Conclusions

1. Within this study

Stable correlation matrices resulted from analyses (A, B, C and D) of 10-year runs of daily data and of 5-daily, monthly and quarterly means. Less stable matrices resulted from 10-year runs of daily data from each quarter (analyses S, T, U, V). The first five components accounted for 81 to 93 per cent of variation. In general, the first component expressed temperature, the second dampness, the third windiness and the fourth snowfall. Some instability resulted in the components from wind direction and speed.

The removal of redundant variables by Jolliffes "B4" method left one temperature variable (dry bulb or maximum), cloud cover, windspeed and direction, rainfall and snow depth as the main sources of variation. These corresponded closely to Lamb's (1972) assessment of the sources of seasonal change in climate, except that visibility as an equivalent of atmospheric transparency, was not retained.

2. For the Study of Stations distributed over the United Kingdom

Economy in the U.K. analysis can be achieved by using data consisting of monthly or quarterly means, since the present study has shown that little change will result in the meanings of the principal components by doing so. This choice will enable fewer values of the principal components to be estimated for each site, and will result in economy in the subsequent analysis relating plant response to the environment. Quarterly means should preferably be calculated on the alternative basis discussed. An analysis of annual means is a possibility which should be examined, but would have been inapplicable for data of one station, because of the close similarity between annual means. Such an analysis would give even greater economy.

The use of separate analyses of quarterly data is not beneficial since this is associated with a reduction in stability, as well as the obvious difficulty that the seasons do not begin and end on single arbitrary dates over the whole United Kingdom.

Transformation of variables does not greatly alter the results when the correlation matrix is in effect calculated from standardised variables.

Reduction in the number of variables has been carried out for this one station, and it would be possible to represent most of the variation present by variables shown in Table 10, which do not exceed 7 in number.

When the factor of geographical variation is added in the U.K. analysis, it is possible that different variables may appear as the main sources of variation, and that reducing the number of variables may not retain the same selection which resulted here. Cost of computing time may make a selection necessary, however, before carrying out a principal component analysis. Though the addition of daylength does not appear important here, it is thought that it may be important to include it in the U.K. analysis.

VI Acknowledgements

The Director-General and the staff of the Meteorological Office are thanked for making the original punched cards available and for commenting on data preparation and analysis, and on aberrant values. Mr. F. H. W. Green and Mr. A. Millar are thanked for discussing various aspects of the study, and the latter also for commenting on aberrant values. Miss L. Goldsmith is thanked for wrting the program CTOP in the PAL-D language, and the program SLCT in 8K Fortran. Mr. J. N. R. Jeffers and Mr. A. H. F. Brown are thanked for making helpful comments on the draft.

VII Summary

Principal Component analyses were carried out on records from the Moor House Climatological Station over a period of ten years, based on daily data, to which the Penman estimate of evapotranspiration was added. Analyses were compared of 10-years'run of data with 5-daily, monthly and quarterly means, and with 10-years'run of daily data divided into quarters and with annual analyses of daily data. The effects of transforming variables and of reducing numbers of variables were also examined. A total of 63 correlation matrices were calculated. Essentially the first four principal components expressed temperature, dampness, windiness and snowfall respectively, and the first five accounted for 81 to 93% of variability. The analysis provides information for a principal component analysis of data from climatological stations scattered over the United Kingdom, for which it forms a pilot study.

VIII References

Anon (1956). Observers' Handbook. Meteorol. Office Publ. 554, HMSO.

Anon (1971). First Report, Royal Commission on Environmental Pollution. HMSO.

Bryson, R. A. and Lahey, J. F. (1958). The March of the Seasons. Univ. Wisconsin, Final Rept., Contract AF 19-(604)-992).

Bunce, R. G. H. (1970). Site classification (a) native woodlands. Merlewood Res. Stn. Rpt. 1967-69, 12-15.

Craddock, J. M. and Flood, C. R. (1969). Eigenvectors for representing the 500 mb. geopotential surface over the northern hemisphere. Q. J. R. met. Soc. 95, 576-93.

Christensen, W. I. Jr. and Bryson, R. A. (1966). An investigation of the potential of component analysis for weather classification. Mon. Weath. Rev. 94, 697-709.

Frank, E. C. and Lee, R. (1966). Potential solar beam irradiation on slopes. U.S. For. Serv. Pap. RM-18.

Fourt, D. R., Donald, D. G. M., Jeffers, J. N. R. and Binns, W. O. (1971). Corsican pine (Pinus nigra var. maritima (Ait.) Melville) in Southern Britain: a study of growth and site factors. Forestry, 44, 189-207.

Fritts, H. C. (1960). Multiple regression analysis of radial growth in individual trees. For. Sci. 6, 334-49.

Green, F. H. W. (1953). A remarkable low humidity. Weather, 8, 182-4.

Green, F. H. W. (1964). A map of average potential water deficit in the British Isles. J. appl. Ecol., 1, 151-8.

Green, F. H. W., Watson, A. and Piccozi, N. (1964). Very low humidity in the East of Scotland in December 1962. Met. Mag. Lond. 94, 225-7.

Hiley, W. E. and Cunliffe, N. (1923). Further observations on the relation of the height growth of trees to meteorological conditions. Ann. appl. Biol. 10, 442-52.

Hocker, H. W. (1956). Certain aspects of climate as related to the distribution of Loblolly pine. Ecology 37, 824-34.

Jeffers, J. N. R. (1967). Two case studies in the application of principal component analysis. Appl. Stat. 16, 225-36.

Jolliffe, I. T. (1972). Discarding variables in a principal component analysis. 1. Artificial data. Appl. Stat. 21, 160-73.

Kutzbach, J. E. (1967). Empirical eigenvectors of sea level pressure, surface temperature and precipitation complexes over North America. J. appl. Meteorol. 6, 791-802.

Lamb, H. H. (1965). Frequency of weather types. Weather 20, 9-12.

Lamb, H. H. (1969). The new look of climatology. Nature 223, 1209-15.

Lamb, H. H. (1971). Problems and practice in longer-range weather and climate forecasting, Aberystwyth Symposia in Agric. Meteorol., No. 14. Weather Forecasting for Agric. and Indust. 17-18 March 1971.

Lamb, H. H. (1972). Climate: Present, Past and Future, Vol. I. Methuen, Lond.

Lorenz, E. N. (1956). Empirical orthogonal functions and statistical weather prediction. Sci. Rept. No. 1., Statistical Forecasting Proj., Dept. Meteorol., M.I.T. Cambridge, Mass, Appendix 1, 29-78.

Lund, I. A. (1963). Map pattern classification by statistical methods. J. appl. Meteorol. 2, 56-65.

Manley, G. (1936). The climate of the northern Pennines: the coldest part of England. Q. Jl R. met. Soc. 62,103-15.

Manley, G. (1943). Further climatological averages for the northern Pennines with a note on topographical effects. Q. Jl. R. met. Soc. 69, 251-61.

Manley, G. (1945). The helm wind of Crossfell, 1937-39. Q. Jl. R. met. Soc. 71, 197-219.

Manley, G. (1952). Climate and the British Scene. Collins, Lond.

Millar, A. (1964a). The effect of temperature and day length on the height growth of birch (Betula pubescens) at 1900 feet in the northern Pennines. J. appl. Ecol. 2, 17-29.

Millar, A. (1964b). Notes on the climate near the upper forest limit in the northern Pennines. Q. Jl. For. 58, 239-46.

Newnham, R. M. (1968). A classification of climate by principal components and its relationship to tree species distribution. For. Sci. 15, 25--64.

Pears, N. V. (1968). Post glacial tree-lines of the Cairngorm Mountains, Scotland. Trans. Soc. Edinb., 40, 361-94.

Pearsall, W. H. and Pennington, W. (1947). Ecological history of the English Lake District. J. Ecol. 34, 137-48.

Perry, A. H. (1970). Filtering climatic anomaly fields using principal components analysis. Trans. Pap. Inst. Brit. Geographers. No. 50, 55-72.

Rosenzweig, M. L. (1968). Net primary productivity of terrestrial communities; prediction from climatological data. Amer. Nat. 102, 67-74.

Seal, H. L. (1964). Multivariate Statistical Analysis for Biologists. Methuen, Lond.

Sellers, W. D. (1965). Physical Climatology. University of Chicago Press.

Sellers, W. D. (1968). Climatology of monthly precipitation patterns in the western United States, 1931-1966. Mon. Weath. Rev. 96, 585-95.

Shorr, B. (1956). Empirical orthogonal functions applied to the prediction of the sea-level pressure field. Sci. Rept. No. 1 Statistical Forecasting Proj. Dept. Meteorol. M.I.T. Cambridge, Mass., appendix 2, 79-143.

Simmons, I. G. (1962). The development of the vegetation on Dartmoor. Ph.D. Thesis, Univ. of Lond.

Smith, L. P. (1967). Potential transpiration. Min. Agric. Fish. Fd. Tech. Bull. 16.

Steiner, D. (1965). A multivariate statistical approach to climatic regionalisation and classification. Tijdschr. K. ned. aadrijksk. Genoot., 82, 329-47.

Stidd, C. K. (1967). The use of eigenvectors for climatic estimates. J. appl. Meteorol. 6, 255-64.

Veitch, L. G. (1970). Forecasting Adelaide's maximum temperature using principal components and multiple linear regression. Melbourne, Bur. Met. Austr. Met. Mag. 18, 1-21.

Wareing, P. F. (1956). Photoperiodism in woody plants. Ann. Rev. Pl. Physiol. 7, 191-214.

White, E. J. (1970). Orthogonalized regressions of height increments on meteorological variables. Aberystwyth Symposia Agric. Meteorol. No. 13. "Research Papers in Forest Meteorology". 18 Mar. 1970. (Ed. J. A. Taylor) Cambrian News, Aberystwth I, 1-12.

Table 1 Data contained on Climatological Land Surface Punched Cards Form 1996

No. on card	Name	Units	Permissible Range on cards	A = absent Moor House	No. of variables used in Moor House analysis
1	Station no.				
2	hour of observation				
3	year				
4	month				
5	Shaw week				
6	day				
7	cloud amount	8ths	0 to 9		1
8	wind direction	degrees	0 to 360		2
9	windspeed	knots	0 to 199		3
10	visibility	code	- 2 to + 9		4
11	present weather	code	00 to 99		
12	dry bulb	°F	0 to 199.9		5
13	wet bulb	°F	0 to 99.9		6
14	vapour pressure	mb	0 to 99.9		7
15	relative humidity	%	0 to 100		8
16	dew point	°F	-99 to + 99		9
17	maximum temperature*	°F	0 to 99		10
18	minimum temperature*	°F	0 to 99		11
19	accumulated temp. above 42°F	day-°F	0 to 99	A	
20	accumulated temp. above 42°F	day-°F	0 to 99	A	
21	grass minimum*	°F	-99 to + 99		12
22	bare soil minimum*	°F	-99 to + 99	A	
23	earth temp. at 4 ins depth	°F	-99 to + 99	A	
24	earth temp. at 8 ins depth	°F	-99 to + 99	A	
25	earth temp. at 1 foot depth	°F	-99 to + 99		13
26	earth temp. at 2 feet depth	°F	-99 to + 99	A	
27	earth temp. at 4 feet depth	°F	-99 to + 99	A	
28	rainfall*	inches	0 to 99.99		14
29	state of ground*	code	00 to 90		
30	total sunshine*	hours	0 to 19.9		15
31	total snow depth	inches	0 to 99		16
32	fresh snow depth*	inches	0 to 99		17
33	run of wind*	miles	0 to 1,999	A	
	Estimated P.E.*	inches	(not on cards)		18

* these variables are daily totals or extremes for the 24 hours ending 09.00 hours G.M.T. Others refer to instantaneous readings.

Table 2 Summary of Variables and from 1960-69 5-daily means (n = 730)

	Variable	Minimum	Mean	Ma imum	S.D.	Coefficient of variation
1	Cloud amount	0.800	6.609	9.00	1.425	0.216
2	Wind direction	32.000	207.497	344.000	61.332	0.296
3	Windspeed	1.600	12.266	28.000	4.842	0.395
4	Visibility	1.200	5.662	8.700	1.273	0.225
5	Dry bulb temp.	18.82	41.492	63.280	9.301	0.224
6	Wet bulb	17.90	40.208	59.040	8.478	0.211
7	V.P.	2.980	8.342	15.580	2.518	0.302
8	R.H. %	50.600	89.961	100.000	7.932	0.088
9	Dew Point	14.600	38.423	56.400	8.151	0.212
10	Maximum temp.	25.800	46.359	70.200	9.833	0.212
11	Minimum temp.	7.300	35.283	54.600	8.373	0.237
12	Grass min.	5.400	31.981	51.200	8.598	0.269
13	Earth temp. at 1 ft.	31.420	43.025	57.500	7.236	0.168
14	Rainfall	0.000	0.210	1.038	0.199	0.948
15	Total sunshine	0.000	2.932	13.740	2.408	0.821
16	Total snow depth	0.000	0.799	16.400	2.059	2.577
17	Fresh snow depth	0.000	0.254	6.800	0.752	2.961
18	P.E.	-0.030	0.037	0.180	0.041	1.108

Table 3 List of Analyses carried out

Correlation Matrices were calculated for all combinations of data and length of period indicated.

The last column gives an alphabetical index which provides a brief method of referring to each analysis in the text.

Data	Period Covered	Correlation Matrix only	Eigen-vectors
Daily	10 years together		A
5-daily	10 years together		B
monthly	10 years together		C
quarterly	10 years together		D
Daily	1960		E
	1961		F
	1962		G
	1963		H
	1964		I
	1965		J
	1966		K
	1967		L
	1968		M
	1969		N
Daily	1st quarter, each year separately	O (1-10)	
	2nd quarter, each year separately	P (1-10)	
	3rd quarter, each year separately	Q (1-10)	
	4th quarter, each year separately	R (1-10)	
Daily	1st quarter, merged 10 years		S
	2nd quarter, merged 10 years		T
	3rd quarter, merged 10 years		U
	4th quarter, merged 10 years		V
Daily, but variables 8, 14, 15, 16, 17, 18 transformed to $\log_e$ (var. + 10.0)	1961		W
Daily, 18 variables daylength: 2 variables for wind direction: less visibility and fresh snow depth	1961		X
Daily temperature variables only	1961		Y
Daily, 3 moisture variables	1961		Z1
Daily, 4 moisture variables	1961		Z2

Table 4 Effects of transforming some Variables, especially on their Coefficients of Variation (1961 daily values)

n = 365

	Variable No.	Name	Original	Calms removed	0-180°N and 0-180°E	Transformation: Cosine	Log$_e$ (var. + 10.0)	1 x
				(n = 352)	0.613			
	2	Wind direction	0.445	0.394	0.745	4.459		
	8	R.H. %	0.1252	-	-	-	0.027	0.61
	14	Rainfall	1.710	-	-	-	0.015	
	15	Sunshine	1.141	-	-	-	0.091	
	16	Total snow depth	3.374	-	-	-	0.075	
	17	Fresh snow depth	5.106	-	-	-	0.027	
Correlation coeff. between consecutive sets	2	Wind direction		0.388	0.440 0 322	0.440		
No. of occasions with change between consecutive sets exceeding half the range: variable 2				26	14 42			

Table 5 Half Matrix of Correlation Coefficients for Daily Values of Meteorological Variables, 1960-69 inclusive

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
2	-0.63*																
3	+0.201**	+0.170**															
4	-0.624**	+0.192**	-0.107**														
5	-0.034	+0.089**	-0.066*	+0.149**													
6	+0.065*	+0.081**	-0.040	+0.065*	+0.936**												
7	+0.160**	+0.044	-0.030	-0.050	+0.902**	+0.954**											
8	+0.558**	-0.094**	+0.163**	-0.556**	-0.391**	-0.239**	-0.002										
9	+0.200**	+0.062*	-0.001	-0.061*	+0.910**	+0.963**	+0.978**	+0.014									
10	-0.124**	+0.088**	-0.141**	+0.220**	+0.952**	+0.931**	+0.847**	-0.430**	+0.846**								
11	+0.168**	+0.113**	+0.044	+0.017	+0.836**	+0.866**	+0.847**	-0.111**	+0.857**	+0.783**							
12	+0.273**	+0.100**	+0.127**	-0.073*	+0.733**	+0.781**	+0.792**	+0.022	+0.803**	+0.662**	+0.898**						
13	-0.040	+0.090**	-0.093**	+0.223**	+0.357**	+0.854**	+0.315**	-0.315**	+0.799**	+0.858**	+0.779**	+0.699**					
14	+0.218**	+0.039	+0.190**	-0.160**	-0.013	+0.021	+0.057	+0.200**	+0.070*	-0.034	+0.051	+0.081**	-0.035				
15	-0.662**	+0.108**	-0.136**	+0.511**	+0.334**	+0.238**	+0.100**	+0.645**	+0.086**	+0.429**	+0.094**	-0.019	+0.298**	-0.285**			
16	-0.006	-0.073*	+0.011	-0.133**	-0.522**	-0.528**	-0.460**	+0.154**	-0.504**	-0.510**	-0.489**	-0.424**	-0.461**	-0.068**	-0.118**		
17	+0.066**	-0.067*	+0.058	-0.140**	-0.550**	-0.353**	-0.309**	+0.117**	-0.333**	-0.364**	-0.295**	-0.233**	-0.304**	+0.064**	-0.107**	+0.607**	
18	-0.331**	+0.125**	-0.014	+0.420**	+0.583**	+0.499**	+0.342**	-0.678**	+0.345**	+0.667**	+0.516**	+0.372**	+0.534**	-0.165**	+0.709**	-0.285**	-0.175**

No. degrees of freedom = no. sets -2 = 3653-2

Value of correlation coefficient significant at 0.05 level of probability = 0.062, shown as *

Value of correlation coefficient significant at 0.01 level of probability = 0.081, shown as **

Table 6 Stability of Matrices of Correlation Coefficients (compared with 0.05 level of probability; counting each correlation coefficient once only; total number = 153)

Correlation Matrices	Stability Class Stable	Partly stable	Unstable
10 years' daily 5-daily, monthly quarterly (4 analyses) (degrees of freedom = 3653-2, 730-2, 120-2, 40-2 respectively) (analyses A B C D)	114	37	2
10 annual analyses (daily data) (365-2 d.f.) (analysis E to N)	98	53	2
10 analyses: 1st quarters (daily data) (90-2 d f.)	42	94	17
10 analyses: 2nd quarters (daily data) (90-2 d.f.)	58	85	10
10 analyses: 3rd quarters (daily data) (90-2 d.f.)	81	67	5
10 analyses: 4th quarters (daily data) (90-2 d.f.)	57	81	15
4 analyses of 10 years' run of 1st 2nd, 3rd, and 4th quarters separately (900-2 d.f.): (analyses O, P, Q R)	66	59	28

Table 7 Stability of Each Variable

The number of instances when the correlation coefficient associated with each variable is stable i.e. all positively significantly, all negatively significantly, or all insignificantly correlated over the analyses indicated at the 0.05 level of probability.

	Possible maximum	Variable No. 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Total
4 analyses of 10 years' daily, 5-daily monthly and quarterly (degrees of freedom = 3653-2, 730-2, 120-2 and 40-2 respectively)	17	9	9	7	10	16	17	13	13	12	14	13	11	17	9	15	14	14	15	228
10 analyses of each year: daily data (365-2 d.f.)	17	10	2	2	9	14	13	14	10	11	14	12	13	15	9	10	11	13	14	196
10 analyses for 1st quarter: daily data (approx. 90-2 d.f.)	17	7	2	1	3	7	7	8	5	8	9	7	6	0	1	3	6	3	1	84
10 analyses for 2nd quarter: daily data (approx. 90-2 d.f.)	17	5	2	2	5	9	8	10	7	9	9	9	6	8	6	8	2	3	8	116
10 analyses for 3rd quarter: daily data (approx. 90-2 d.f.)	17	8	3	3	11	8	11	10	10	9	9	9	9	9	2	8	16	17	10	162
10 analyses for 4th quarter: daily data (approx. 90-2 d.f.)	17	6	6	0	5	8	8	9	6	10	9	3	9	11	2	5	6	2	4	114
4 analyses of 10 years' 1st, 2nd, 3rd and 4th quarters (approx. 900-2 d.f.)	17	8	3	7	11	7	7	8	9	9	9	9	13	9	6	8	1	2	6	132
Total scorings	119	53	27	22	54	69	71	72	60	68	73	67	67	69	35	57	56	54	58	

Correlation coefficients are counted twice (i.e. the coefficient for the correlation of variables 6 and 11 is counted under variables 6 and 7). Cases are underlined with less than 25% possible instances in the Stable Class.

Table 8 Eigenvalues and Standardised Eigenvectors for the First Five Principal Components of Ten-Yearly Analyses of Daily Data, and 5-Daily, Monthly and Quarterly Means.

No.	Variable	1st Principal Component Data				2nd Principal Component Data			
		Daily	5-daily	Monthly	Quarterly	Daily	5-daily	Monthly	Quarterl
1	Cloud	-0.033	-0.165	-0.246	-0.364	1.000	1.000	1.000	0.943
2	Wind direction	0.127	0.177	0.335	0.364	-0.134	0.032	0.416	0.686
3	Windspeed	-0.055	-0.161	-0.273	-0.471	0.300	0.377	0.418	0.229
4	Visibility	0.192	0.427	0.713	0.806	-0.880	-0.749	-0.526	-0.357
5	Dry bulb	1.000	1.000	0.995	0.995	-0.002	0.087	0.129	0.136
6	Wet bulb	0.998	0.989	0.985	0.987	0.162	0.199	0.205	0.202
7	V.P.	0.931	0.942	0.955	0.955	0.367	0.326	0.277	0.297
8	R.H.	-0.345	-0.583	-0.777	-0.826	0.935	0.807	0.676	0.580
9	D.P.	0.937	0.945	0.960	0.966	0.396	0.366	0.320	0.303
10	Max. temp.	0.981	1.000	1.000	1.000	-0.141	-0.003	0.065	0.089
11	Min. temp.	0.918	0.909	0.921	0.956	0.290	0.356	0.277	0.247
12	Grass min.	0.822	0.856	0.926	0.934	-0.445	0.525	0.393	0.379
13	Soil temp. 1 ft.	0.927	0.941	0.958	0.961	-0.033	0.102	0.163	0.263
14	Rain	-0.031	-0.112	-0.180	-0.269	0.449	0.638	0.779	1.000
15	Sunshine	0.356	0.501	0.767	0.802	-0.949	-0.829	-0.690	-0.602
16	Total snow	-0.614	-0.585	-0.669	-0.765	-0.030	-0.129	-0.305	-0.155
17	Fresh snow	-0.435	-0.524	-0.638	-0.703	0.049	-0.019	-0.315	-0.136
18	P.E.	0.657	0.802	0.863	0.892	-0.643	-0.406	-0.412	-0.374
	Eigenvalues	7.93	9.15	10.83	11.75	3.42	3.12	2.34	2.34
	% variability	44.06	50.84	60.16	65.30	19.01	17.33	13.00	13.00
	Cumulative	44.06	50.84	60.16	65.30	63.07	68.17	73.16	78.30

(Eigenvectors over +0.7 or less than -0.7 are underlined)

Table 8 cont'd

3rd Principal Component Data				4th Principal Component Data				5th Principal Component Data			
Daily	5-daily	Monthly	Quarterly	Daily	5-daily	Monthly	Quarterly	Daily	5-daily	Monthly	Quarterly
0.024	-0.131	-0.441	0.315	0.031	0.074	0.244	0.167	0.193	0.116	0.288	0.311
0.880	1.000	1.000	0.649	-0.499	-0.289	0.055	-0.530	0.250	-0.739	-0.491	-0.814
1.000	0.919	0.925	0.122	-0.299	-0.106	-0.028	-1.000	0.262	1.000	1.000	1.000
0.205	0.454	0.558	-0.183	-0.227	-0.214	-0.194	-0.395	-0.188	-0.198	-0.144	-0.394
-0.013	-0.039	-0.069	0.083	0.121	0.106	0.116	0.080	-0.043	-0.016	0.048	0.116
-0.030	-0.055	-0.095	0.059	0.102	0.090	0.121	0.104	-0.027	-0.043	0.044	0.128
-0.080	-0.108	-0.158	0.047	0.138	0.135	0.196	0.197	-0.01[illegible]	-0.110	0.028	0.14[illegible]
-0.109	-0.135	-0.208	-0.322	-0.103	-0.137	-0.046	0.171	0.103	-0.171	-0.0[illegible]4	0.04
-0.057	-0.083	-0.135	0.01[illegible]	0.076	0.072	0.126	0.13[illegible]	-0.002	-0.0[illegible]	0.038	0.[illegible]42
-0.055	-0.040	-0.045	0.081	0.127	0.093	0.096	0.053	-0.031	-0.025	0.02	[illegible]
0.134	0.046	0.006	0.113	0.074	0.100	0.136	0.031	0.075	0.031	0.056	0.[illegible]2
0.193	0.042	-0.14	0.098	0.075	0.127	0.202	0.18	0.140	0.052	0.[illegible]3	0.226
-0.006	-0.03	-0.146	0.012	0.141	0.156	0.226	0.230	-0.022	-0.141	-0.049	0.042
0.589	0.709	0.7[illegible]	-0.144	0.013	0.142	0.249	-0.185	-1.000	-0.186	-0.421	-0.08[illegible]
0.096	0.127	0.262	0.358	0.158	0.170	0.056	-0.239	0.0[illegible]6	0.217	0.182	-0.103
0.320	0.165	0.130	0.829	0.809	0.911	0.980	0.25[illegible]	-0.020	-0.221	0.034	-0.012
0.445	0.185	0.173	1.000	1.000	1.000	1.000	0.29[illegible]	0.136	-0.065	0.162	-0.13[illegible]
0.286	0.186	0.178	0.381	0.190	0.254	0.[illegible]4	-0.196	0.112	0.393	0.223	0.220
1.28	1.49	1.26	1.04	1.21	1.26	1.09	0.94	0.86	0.[illegible]0	0.[illegible]3	0.69
7.13	8.30	6.99	5.80	6.69	7.00	6.04	5.24	4.78	3.86	4.08	3.82
70.20	76.47	80.15	84.10	6.89	83.47	86.19	89.34	81.67	87.33	90.27	93.16

Table 9 Eigenvalues and standardised eigenvectors for 10 years' analyses of daily data in quarters

Variable No.	1st Principal Component Quarter 1	1st PC 2	1st PC 3	1st PC 4	2nd Principal Component 1	2nd PC 2	2nd PC 3	2nd PC 4	3rd Principal 1	3rd Principal 2
1	0.459	-0.348	0.394	0.372	0.977	1.000	0.946	0.972	0.281	-0.182
2	0.200	0.036	-0.216	-0.001	-0.392	-0.127	-0.053	-0.126	-0.202	-0.782
3	0.280	-0.180	-0.039	0.160	0.141	0.204	0.318	0.379	0.725	-0.845
4	-0.524	0.324	-0.056	-0.261	-0.933	-0.860	-0.655	-0.939	-0.158	-0.432
5	0.972	1.000	0.679	0.988	-0.233	-0.111	-0.891	-0.080	0.006	0.209
6	1.000	1.000	0.921	1.000	0.106	0.182	-0.515	0.032	-0.078	0.244
7	0.992	0.836	1.000	0.980	0.090	0.593	-0.080	0.154	-0.251	0.352
8	0.336	-0.453	0.507	0.175	0.892	0.991	0.959	0.955	-0.532	0.117
9	0.995	0.842	0.992	0.973	0.136	0.621	-0.043	0.220	-0.218	0.293
10	0.797	0.983	0.578	0.894	-0.57	-0.185	-0.087	-0.326	0.022	0.212
11	0.724	0.784	0.838	0.894	-0.149	0.643	-0.015	-0.139	0.36[illegible]	-0.111
12	0.831	0.543	0.755	0.8 4	-0.008	0.771	0.156	0.087	0.505	-0.159
13	0.428	0.885	0.514	0.827	-0.611	0.152	-0.555	-0.402	0.099	0.020
14	0.209	-0.179	0.150	0.089	0.348	0.485	0.435	0.434	0.600	-0.402
15	-0.345	0.451	-0.368	-0.033	-1.000	-0.96	-1.000	-1.000	-0.076	0.108
16	-0.648	-0.553	-0.008	-0.500	0.494	-0.223	0.017	0.399	0.388	0.996
17	-0.543	-0.444	-0.008	-0.357	0.511	-0.071	0.011	0.318	0.891	1.000
18	-0.023	0.673	-0.242	0.252	-0.860	-0.715	-0.965	-0.720	1.000	-0.248
Eigenvalue	7.03	7.10	5.59	7.27	3.13	4.06	4.09	3.20	1.39	1.48
% Variability	39.06	39.44	31.03	40.41	17.39	22.53	22.73	17.78	7.71	8.20
Cumulative %	39.06	39.44	31.03	40.41	56.45	61.97	53.76	58.19	64.16	70.17

(Eigenvectors over +0.7 or less than -0.7 are underlined)

Table 9 cont'd

Component		4th Principal Component				5th Principal Component			
3	4	1	2	3	4	1	2	3	4
-0.005	-0.070	0.255	-0.020	-0.099	0.018	0.285		0.1[illegible]3	0.454
-0.003	1.000	-1.000	-0.727	-0.786	-0.739	0.364		0.1[illegible]4	0.268
-0.003	0.816	-0.542	-0.989	-1.000	-0.835	1.000		-0.175	0.00[illegible]
-0.002	0.244	-0.374	-0.081	-0.119	-0.300	-0.020		-0.212	-0.138
0.005	0.023	-0.018	0.059	0.167	0.028	-0.314		-0.123	-0.216
0.005	-0.009	-0.040	0.063	0.170	0.017	-0.271		-0.065	-0.238
0.006	-0.067	-0.092	0.035	0.146	0.056	-0.192		0.003	-0.322
-0.003	-0.217	-0.121	-0.020	0.012	-0.007	0.256		0.138	-0.235
0.006	-0.061	-0.072	0.056	0.154	-0.001	-0.185		0.002	-0.282
0.008	0.001	-0.030	0.091	0.151	0.213	-0.539		0.008	-0.247
0.004	0.272	0.142	-0.400	-0.640	0.144	-0.032		0.065	0.465
0.003	0.258	0.171	-0.623	-0.726	0.066	0.085		0.101	0.223
-0.018	0.084	0.297	-0.242	-0.025	0.264	0.780		-0.056	-0.049
-0.004	0.733	-0.662	-0.062	-0.132	-0.023	-0.573		-1.000	0.005
0.004	0.083	-0.066	-0.220	-0.164	0.208	-0.250		0.055	-0.582
1.000	0.733	-0.396	-0.792	-0.002	0.723	-0.606		-0.002	-0.318
1.000	0.742	-0.100	-1.000	-0.002	1.000	-0.306		-0.002	-0.234
0.002	0.511	0.346	-0.532	-0.692	0.356	0.018		0.042	1.000
2.00	1.53	1.12	1.35	1.60	1.13	1.03		0.94	0.96
11.11	8.53	6.20	7.51	8.87	6.27	5.74		5.22	5.34
64.87	66.72	70.36	77.68	73.74	72.99	76.10		78.96	78.33

Convergence was not achieved in 30 mins. for the 5th principal component of the 2nd quarter.

Table 10 Results of Reduction of variables by Jolliffe's "B4" Method, showing variables retained in descending order of importance

10-year analyses				1961 daily data with daylength and 2 variables for wind direction; without visibility and fresh snow depth (analysis X)
Daily data	5-daily means	Monthly	Quarterly	
dry bulb temp.	dry bulb temp.	max. temp.	max. temp.	dry bulb temp.
cloud	cloud	cloud	rain	cloud
windspeed	wind direction	wind direction	fresh snow	windspeed
fresh snow	fresh snow	fresh snow	windspeed	wind direction ^{o}N
rain	windspeed	windspeed	-	rainfall
wind direction	-	-	-	wind direction ^{o}E
-	-	-	-	total snow depth
Value of last eigenvalue above 0.71				
0.81	0.70*	0.73	0.94	0.75

(*For the 5-day means the next lower value has been taken since the value above was considerably above i.e. 1.26)

Table 11   Quarterly values of Empirical Orthogonal Functions Derived from First Five Principal Components of Ten-yearly analyses of Quarterly Data

		1	2	3	4	5
1960	1	-4.021	-0.934	-0.763	0.872	0.836
	2	3.927	-3.160	0.053	-0.506	-0.255
	3	4.520	-0.098	-0.140	1.398	-0.50[illegible]
	4	-1.857	1.312	-2.044	1.350	-0.773
1961	1	-3.459	0.480	0.495	-1.050	0.799
	2	2.526	-1.532	0.072	-0.350	0.368
	3	4.594	0.526	-0.006	-0.493	-0.480
	4	-1.602	-0.009	-1.711	0.314	-1.084
1962	1	-4.095	-0.103	-0.579	-1.377	-0.519
	2	2.021	-1.292	0.482	-1.277	-0.549
	3	3.381	1.558	0.084	-0.957	0.369
	4	-1.349	-0.896	-1.197	-0.661	-0.211
1963	1	-5.145	-3.887	-0.254	0.625	-0.395
	2	1.556	-0.465	-0.267	-0.868	0.847
	3	3.833	0.301	-0.181	-0.288	0.164
	4	-1.911	1.221	-1.397	-0.131	0.956
1964	1	-3.630	-1.285	-1.520	0.776	1.591
	2	2.041	-0.447	0.428	1.326	1.683
	3	4.110	0.906	0.471	-0.797	1.076
	4	-1.611	-0.258	-1.384	0.350	-0.977
1965	1	-4.530	-1.250	0.460	0.059	0.104
	2	2.187	-1.078	0.326	-0.758	-0.221
	3	3.103	1.960	0.362	0.528	0.668
	4	-3.145	0.796	1.441	0.522	-0.064
1966	1	-3.950	1.483	-0.345	-1.104	1.525
	2	1.327	-0.50	0.538	0.219	0.504
	3	4.395	0.541	0.488	0.747	-0.683
	4	-2.119	2.13	-0.191	0.001	-0.830
1967	1	-3.471	1.856	0.874	-1.753	-0.145
	2	1.655	-0.361	0.529	-0.226	-0.112
	3	4.315	2.118	0.454	1.428	-0.634
	4	-1.845	2.367	0.189	-1.569	-1.810
1968	1	-4.585	1.049	2.190	0.572	-1.220
	2	2.511	-3.225	-0.264	-0.181	-0.943
	3	3.672	1.410	-0.239	1.797	0.669
	4	-2.555	1.492	-1.362	1.502	0.723
1969	1	-6.301	-1.063	3.478	1.762	0.511
	2	1.993	-2.327	0.084	0.133	-0.536
	3	5.128	-0.150	0.715	1.136	0.484
	4	-1.614	0.900	-0.370	-0.319	-0.929

Table 12a Eigenvalues and Standarised Eigenvectors of Six Variables expressing temperature (Daily Values for 1961) for the First Four Principal Components (Analysis Y)

Variable	Component			
	1	2	3	4
Dry bulb	0.997	0.325	0.418	0.379
Wet bulb	1.000	0.165	0.444	0.540
Max.	0.954	0.659	0.140	-1.000
Min.	0.970	-0.591	0.024	0.127
Grass min.	0.901	-1.000	-0.096	-0.406
Earth at 1 ft.	0.930	0.384	-1.000	0.300
Eigenvalue	5.21	0.43	0.21	0.08
% Variability	86.95	[illegible].19	3.43	1.31
Cumulative %	86.95	94.14	97.57	98.88

Correlation coefficient between dry bulb and wet bulb 0.986)
Correlation coefficient between wet bulb and grass min. 0.806) Sign. at 0.001 level of probability
Correlation coefficient between dry bulb and grass min. 0.773)

Table 12b Eigvenvalues and standardised eigenvectors for the first three components of variables representing airborne moisture at screen height calculated from daily data for 1961 (analysis Z1)

Variable	Component		
	1	2	3
Vapour pressure (mb)	1.000	0.004	1.000
Relative humidity (%)	0.000	-1.000	0.009
Dew point (^{O}F)	1.000	-0.004	-1.000
Eigenvalue	1.97	1.00	0.03
% Variability	65.68	33.33	0.99
Cumulative %	65.68	99.01	100.00

Table 13 Interpretation of Principal Components of Meteorological Variables in Relevent Studies (Cumulative percentages accounted for are shown)

	Study				
	Present paper	White (1970)	Fourt et al (1971)	Newnham (1968)	Christensen and Bryson (1966)
No. variables	13	12	15	19	15
periodicity	daily (essentially)	several days	10 yearly seasonal and monthly means	quarterly	twice daily
derived variables	P.E.	smoothed P.E.	P.E.	length of frost-free period	u and v components from wind direction and speed
non-meteorological variables	none	none	Aspect altitude exposure, latitude longitude	latitude altitude	none
period covered	10 years (1960-69)	1 growing season	10 years (1957-66)	mostly 14 years	4 years of January 5 years of July
First principal Component	temperature (44.1)	energy (35.4)	energy balance (50.0)	general index of winter and fall climate and length of growing season (precipitation and temperature) (57.4)	temperature (16.3)
Second component	dampness contrasted with dryness (63.1)	soil temperature (59.4)	summer warmth and winter cold (78.9)	contrast between spring and summer mean and max. temp. and precipitation (86.8)	unique combination of wind and pressure (shearing stress unimportant) (34.5)
Third component	wind direction and speed (70.2)	storminess (73.7)	aspect (85.0)	latitude (92.2)	
Fourth component	snowfall (76.9)	mean P.E. (80.8)	P.E. in June (90.4)	spring and summer rainfall (95.8)	cloud cover at noon (43.4)
Fifth component	rainfall (81.7)	rainfall (87.1)	-	-	synoptic (55.5)

Figure 1 Flow diagram showing method of correcting large matrices of data on punched paper tape (Computer program names in capitals)

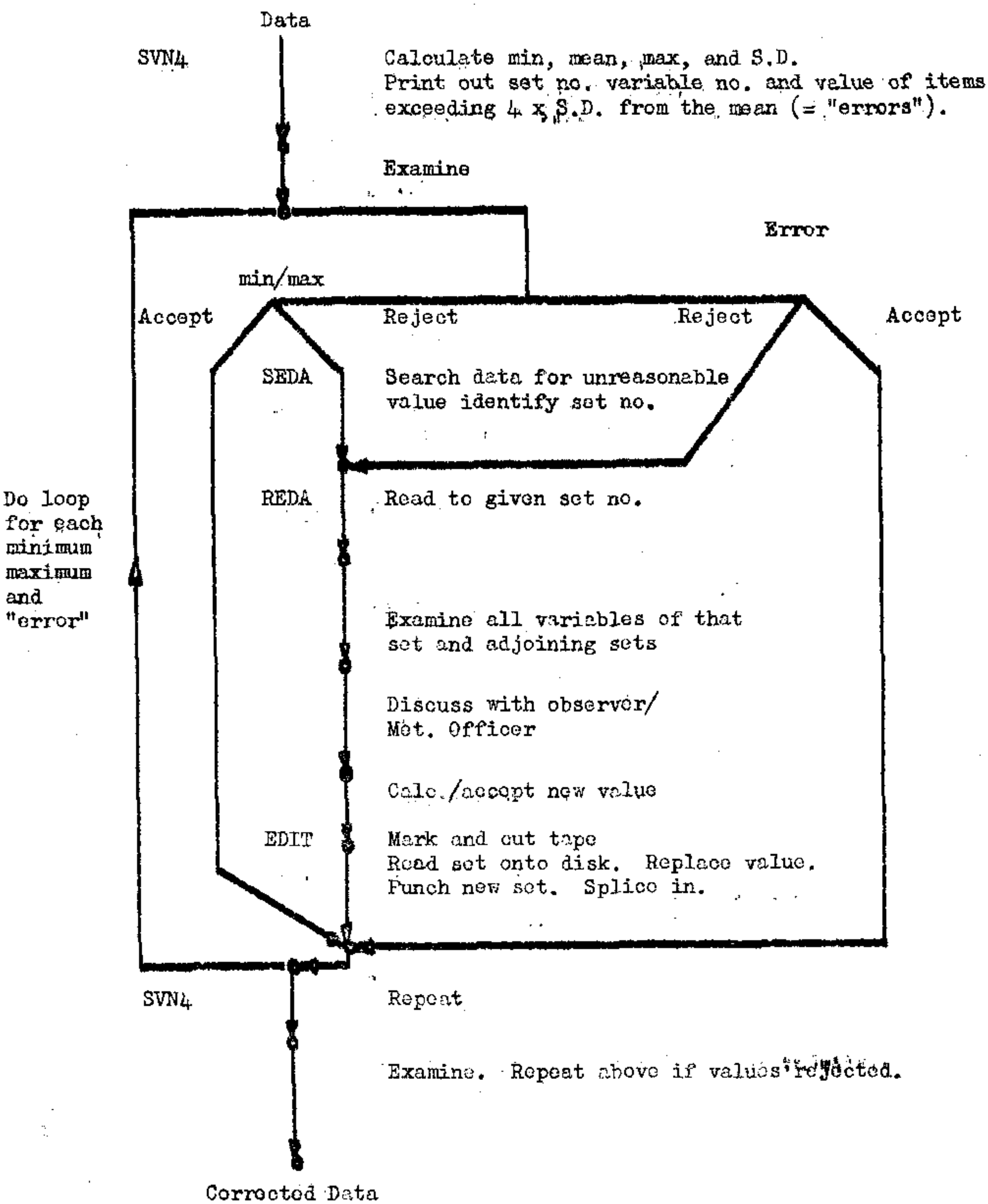

FIG. 2 FLOW DIAGRAM SHOWING PREPARATION OF DATA PRIOR TO ANALYSIS

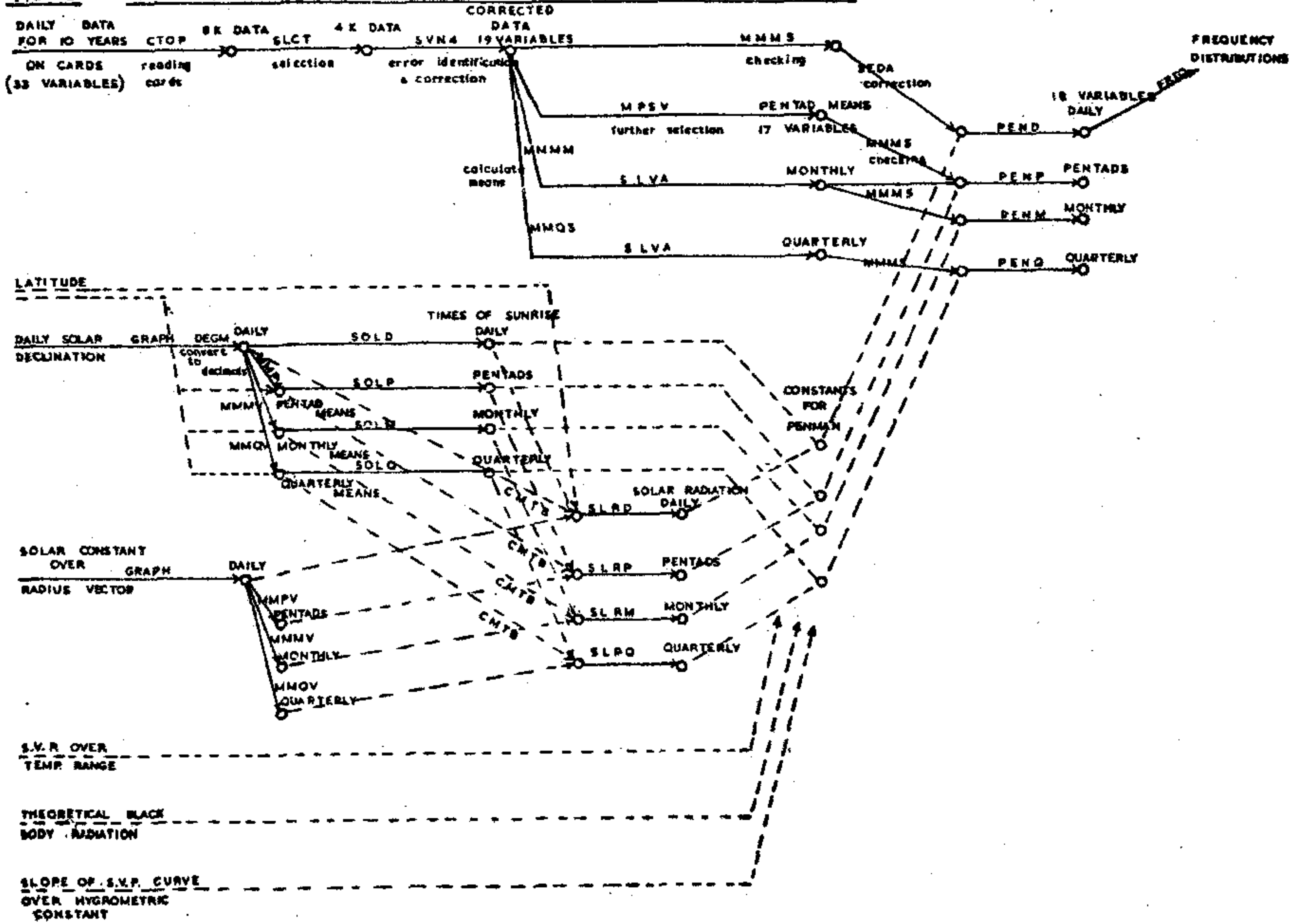

FIG. 3 FLOW DIAGRAM SHOWING ANALYSIS METHODS INVOLVING 61 CORRELATION MATRICES

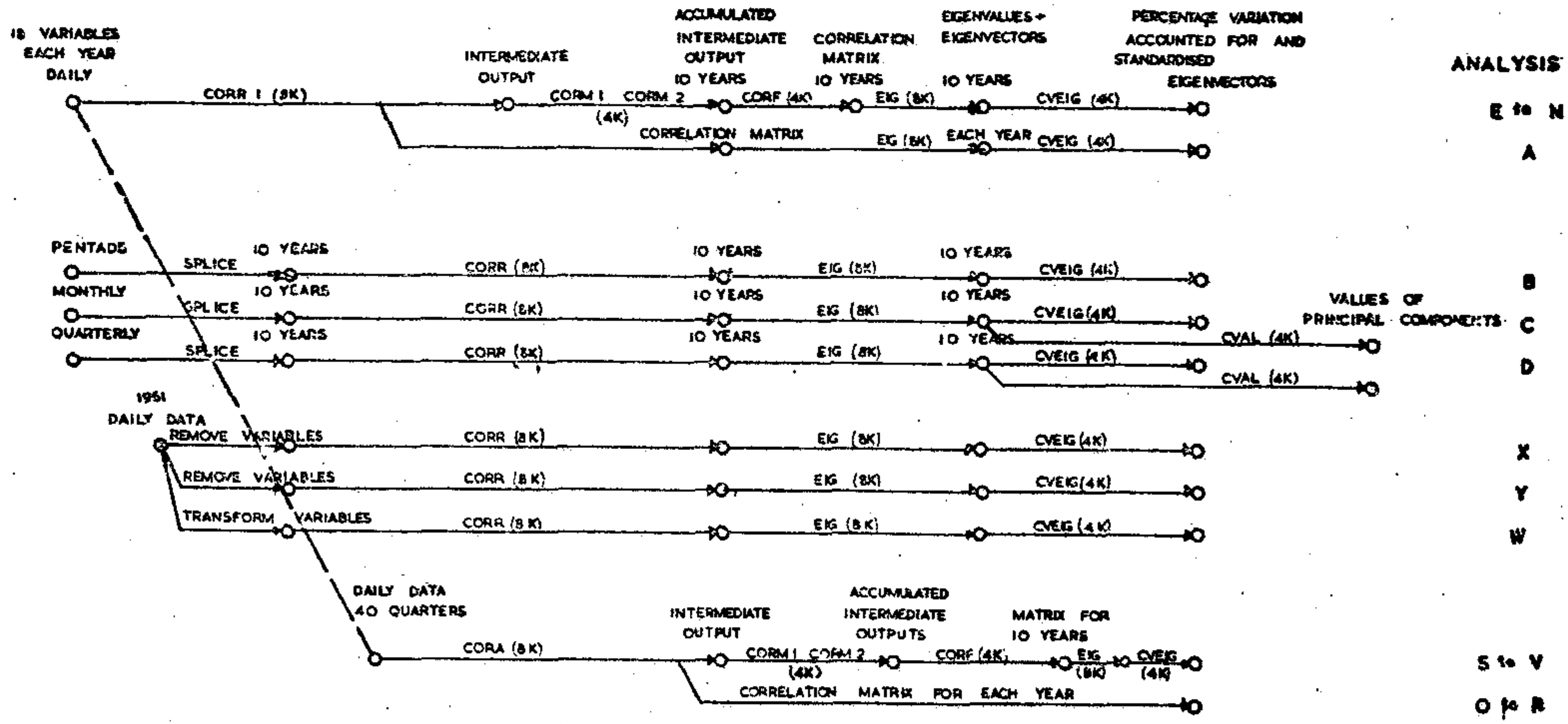

FIG. 4 FREQUENCY DISTRIBUTIONS OF 1961 DAILY DATA

(RANGE OF EACH DIVIDED INTO 10 CLASSES)

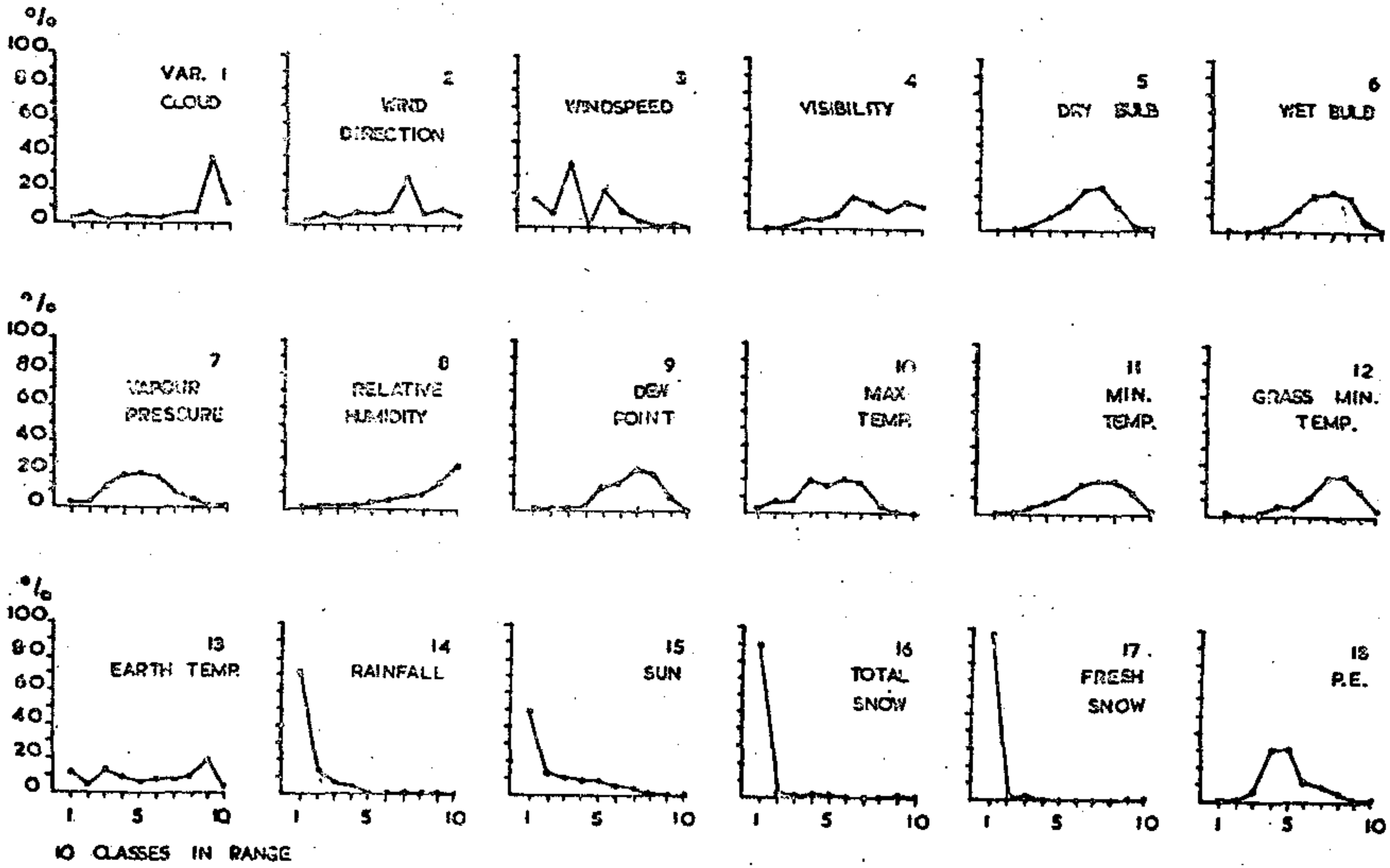

FIG. 5

REPRESENTATIVE SCATTER DIAGRAMS SHOWING THE VARIATION FOUND IN CORRELATION COEFFICIENTS OVER THE ANALYSES CARRIED OUT

VALUES OF CORRELATION COEFFICIENTS INDICATING SIGNIFICANCE AT THE 0·05 LEVEL OF PROBABILITY ARE SHOWN.

STABILITY CLASSES OF STABLE, PARTLY STABLE AND UNSTABLE ARE INDICATED BY THE LETTERS S, P AND U.

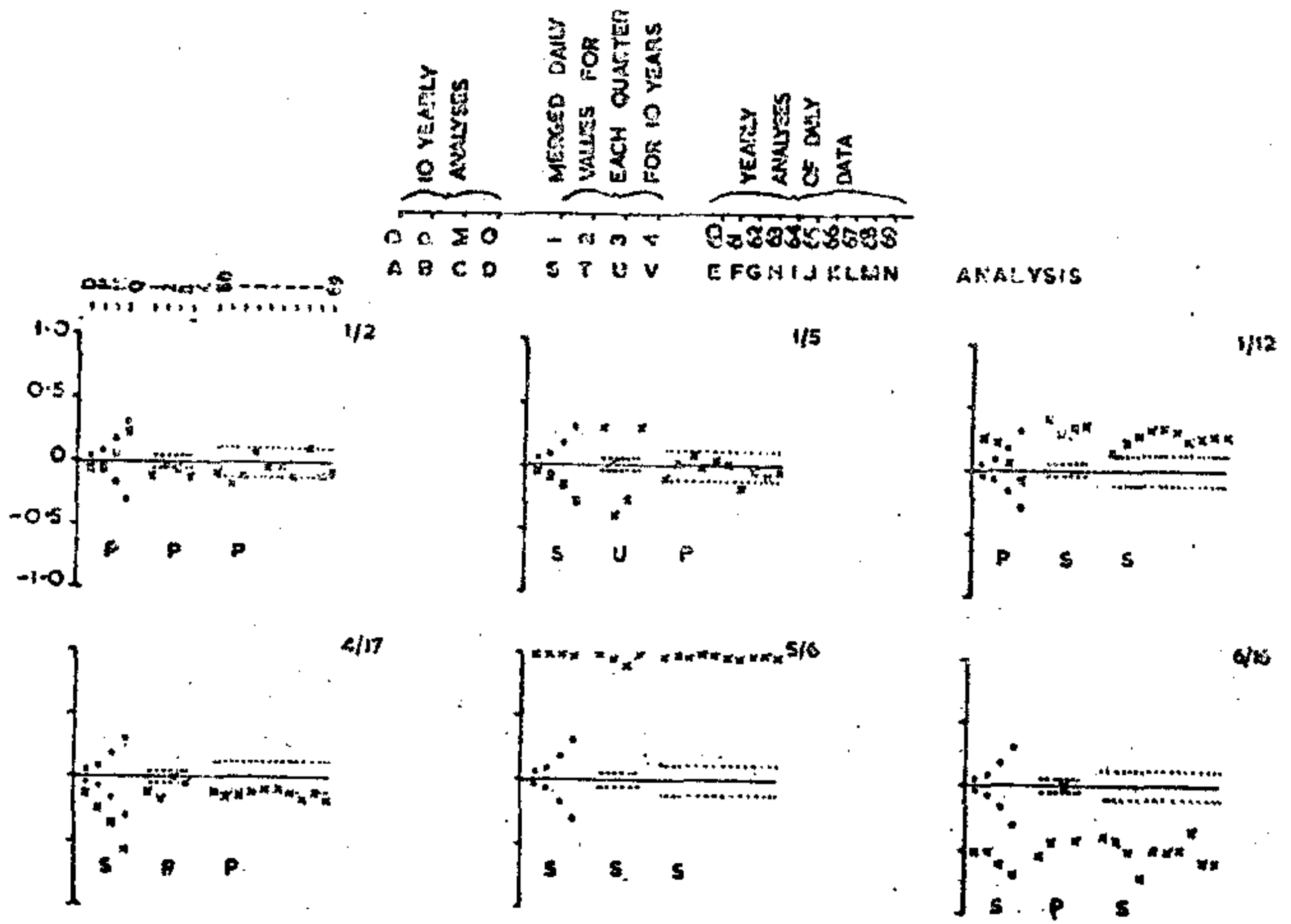

FIG. 6.

REPRESENTATIVE SCATTER DIAGRAMS SHOWING THE VARIATION FOUND IN CORRELATION COEFFICIENTS OVER FOUR ANALYSES OF TEN YEARS' DAILY DATA OF FIRST, SECOND, THIRD AND FOURTH QUARTERS.

VALUES OF CORRELATION COEFFICIENTS INDICATING SIGNIFICANCE AT THE 0·05 LEVEL OF PROBABILITY ARE SHOWN.

STABILITY CLASSES OF STABLE, PARTLY STABLE AND UNSTABLE ARE INDICATED BY THE LETTERS S, P AND U.

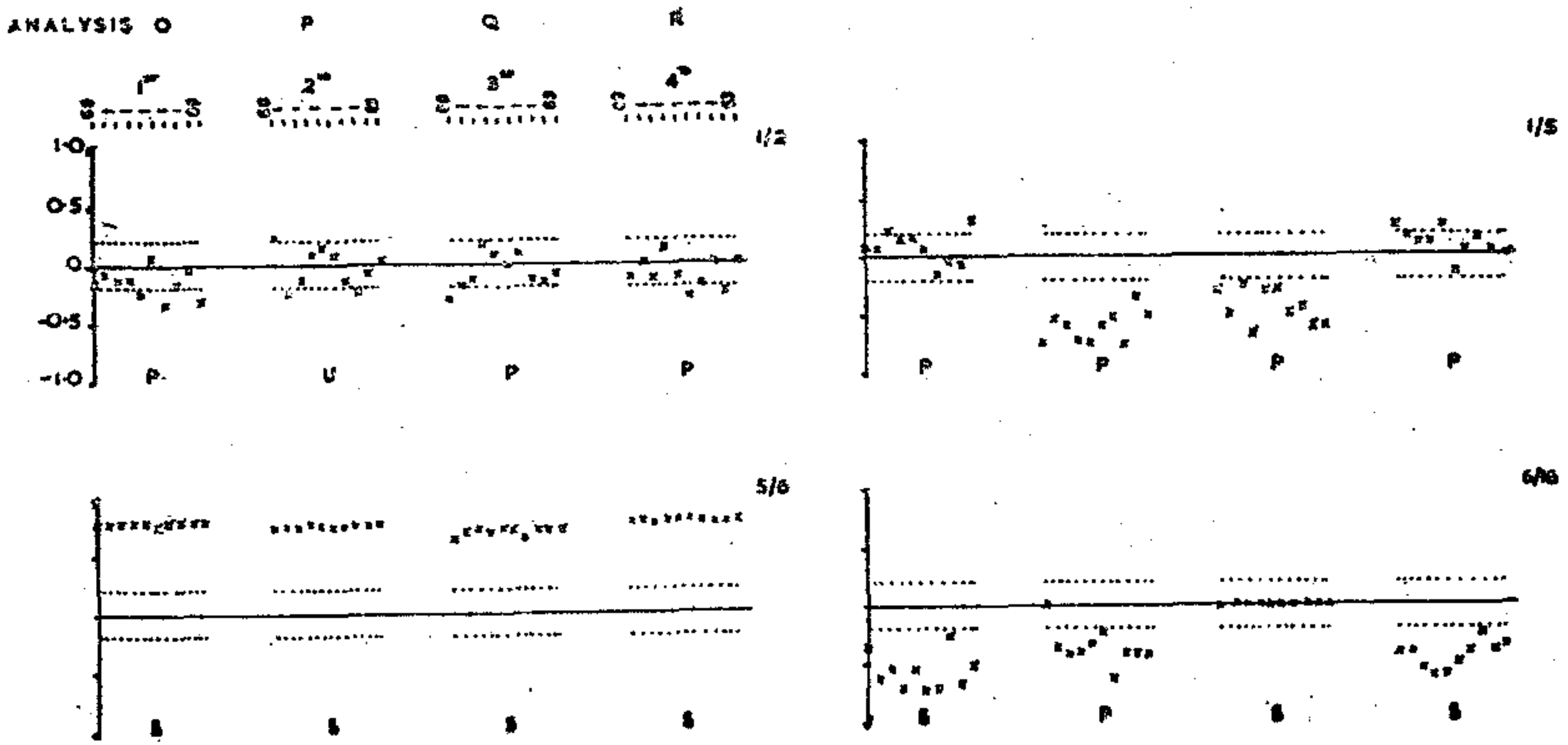

FIG.7

VARIATION ACCOUNTED FOR BY THE FIRST FIVE PRINCIPAL COMPONENTS.

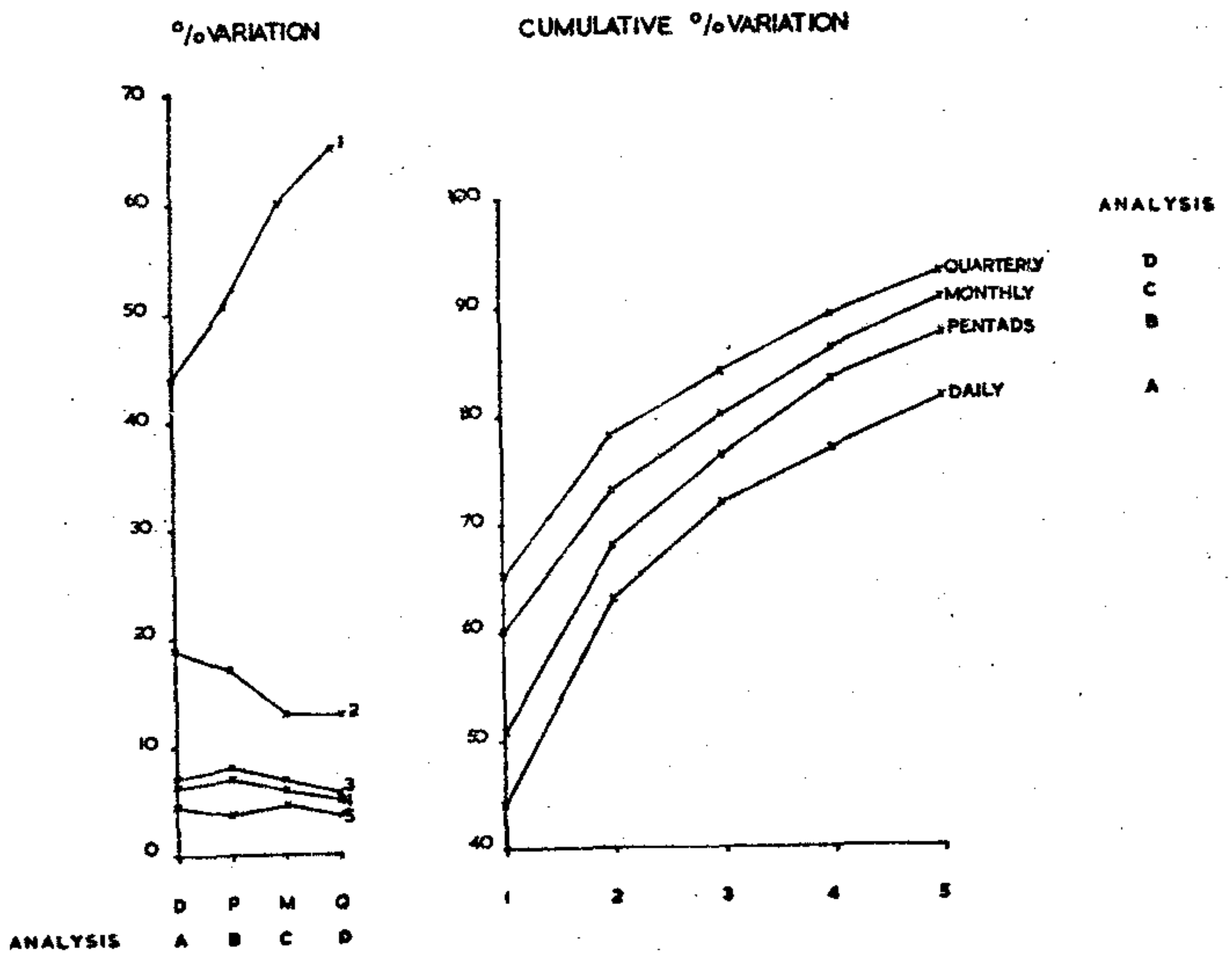

FIG. 8

STANDARDIZED VALUES OF EIGENVECTORS OF THE FIRST PRINCIPAL COMPONENT

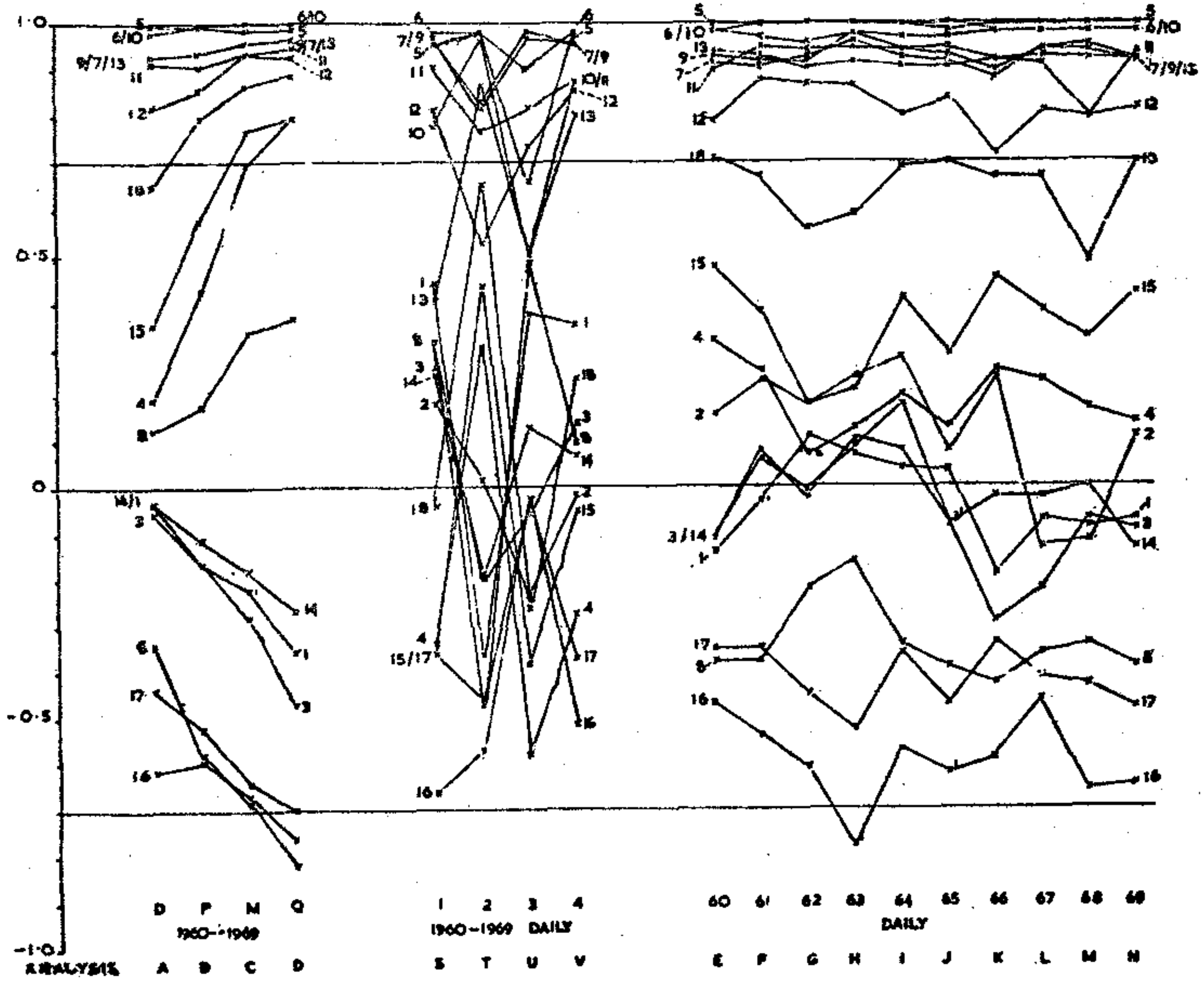

FIG. 9

QUARTERLY VALUES OF PRINCIPAL COMPONENTS

(CALCULATED FROM ANALYSIS D)

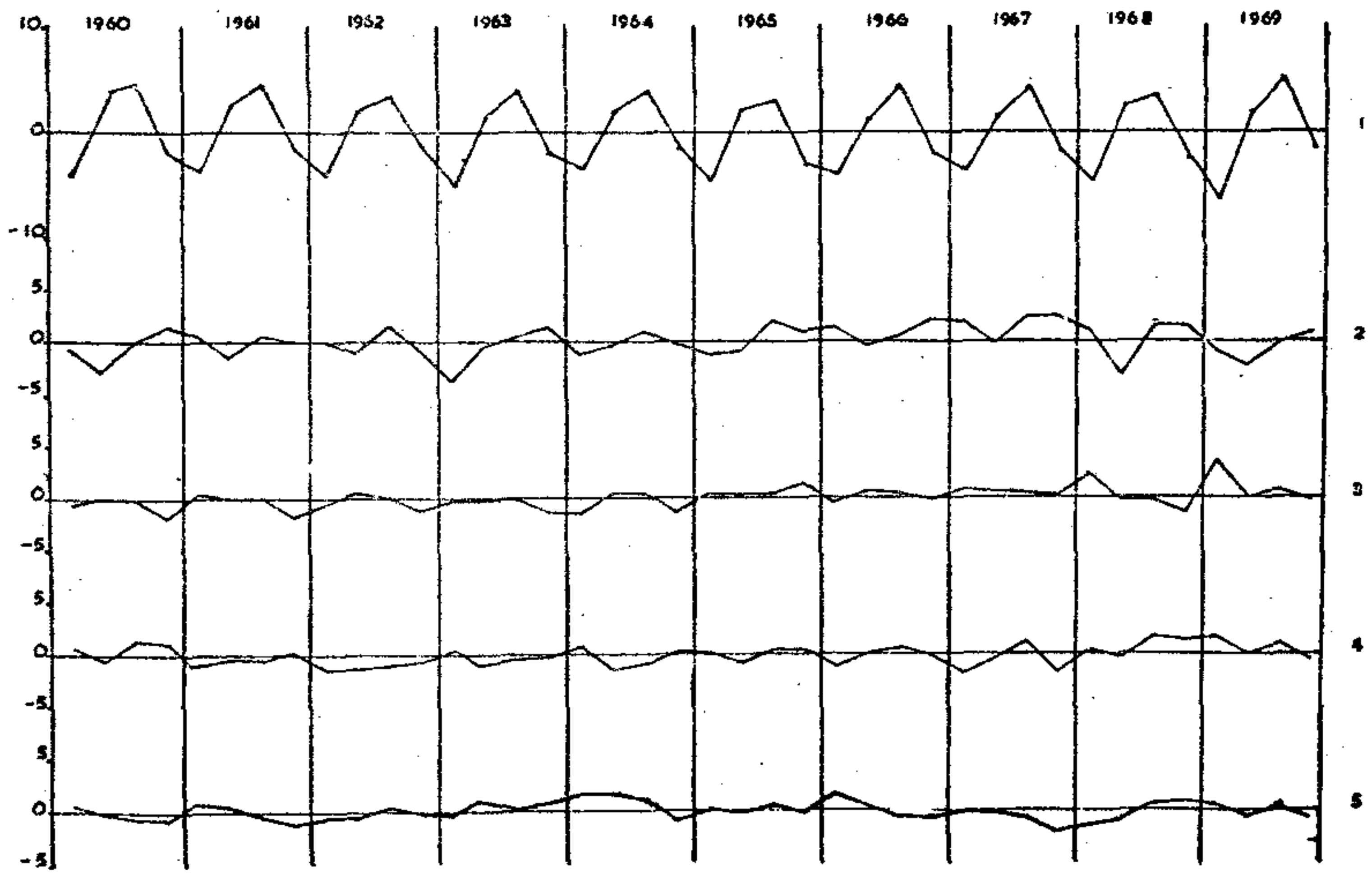

FIG. 10

MONTHLY VALUES OF PRINCIPAL COMPONENTS.

(CALCULATED FROM ANALYSIS C)

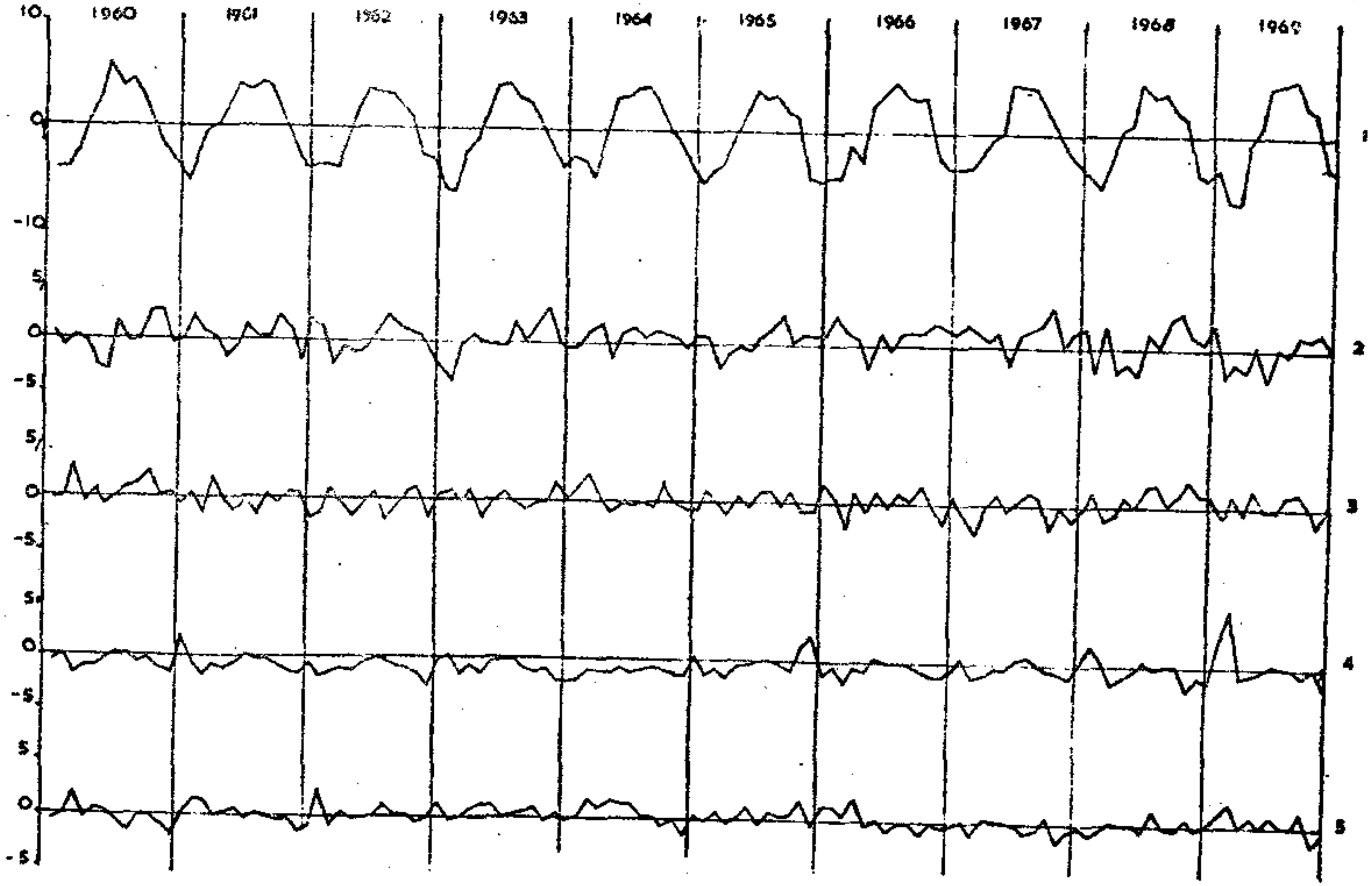

FIG 11 MEAN MONTHLY VALUES OF FIRST FIVE PRINCIPAL COMPONENTS OF METEOROLOGICAL VARIABLES : MOOR HOUSE 1960–69 INCL.