Principal Components Analysis of Vector Wind Measurements[1]

DONALD M. HARDY

Solar Energy Research Institute, Golden, CO 80401

JOHN J. WALTON

Lawrence Livermore Laboratory, Livermore, CA 94550

(Manuscript received 6 July 1977, in final form 20 February 1978)

ABSTRACT

The method of principal components analysis (also known as empirical eigenvector analysis) was generalized to the treatment of vector fields of data and applied to a 12-month record of mean hourly wind velocities from 10 measurement locations in a mesoscale region. The primary spatial distributions of regional wind velocities were derived for each month. Time-series analysis in terms of the primary spatial velocity patterns was used to determine the fundamental temporal patterns or principal components. Necessary mathematical procedures are given and geometric representations of eigenvectors that define the primary spatial velocity patterns are presented. Applications of the generalized vector formulation of the method to current and future problems of atmospheric science are discussed.

1. Introduction

The statistical method of principal components analysis, frequently referred to as empirical eigenvector analysis, and the closely related procedures of factor analysis have been used to study many different types of observational data. Previous studies, however, were limited to the analysis of patterns in scalar fields of data. In the present work, the method has been generalized to the treatment of vector measurements. Specifically, the observed wind velocities in a mesoscale region were treated by a generalized form of principal components analysis. The primary mesoscale spatial velocity patterns were derived, and the method of time-series analysis in terms of the primary spatial patterns was used to determine the essential temporal patterns given by the principal components.

Many applications of principal components analysis to scalar data have been reported in climatic, environmental and meteorological studies. Stidd (1967) applied these methods to derive precipitation patterns in Nevada, noting many other potential applications. Clark and Peterson (1973) studied temporal and spatial variations of the St. Louis heat island by these techniques. A similar analysis of air pollution patterns and their relation to meteorological variables was reported by Peterson (1970). Related factor analysis techniques were used by Blifford and Meeker (1967) to examine the pollutant problems of 30 U.S. cities. Monthly patterns of surface pressure, temperature and rainfall were studied by Kidson (1975) using the method of principal components analysis. A simplified scalar formulation of the present method, as well as numerical examples based on analytic data sets was given by Hardy (1976). A useful summary of principal components analysis and of the formal procedures of empirical eigenvector analysis as previously used in atmospheric science is given by Essenwanger (1976). In addition, King (1969) has written a valuable general discussion of both factor analysis and principal components analysis as they relate to modern geography.

Vector parameters are very important in many physical problems. The present paper describes a new approach in which principal components analysis is used to study two-dimensional vector observations. Treatment of vector observations is achieved by an appropriate use of complex rather than real numbers in the analysis. A brief statement of the necessary mathematical formulation is given by Hardy (1977a). Both the use of complex numbers and the treatment of vector observations represent new concepts in using the method to study physical variables. The ability to study vector observations opens important new possibilities in the physical sciences for this well-known method of data analysis.

All applications of principal components analysis begin with an array of observations or measurements that can be grouped into a rectangular matrix. In most cases, a physical parameter is measured many times at a number of specified geographic locations and ordered by measurement location and time to

[1] This work was performed under the auspices of the U. S. Department of Energy under Contracts W-7405-Eng-48 and EG-77-C-01-4042.

0021-8952/78/1153-1162$05.00

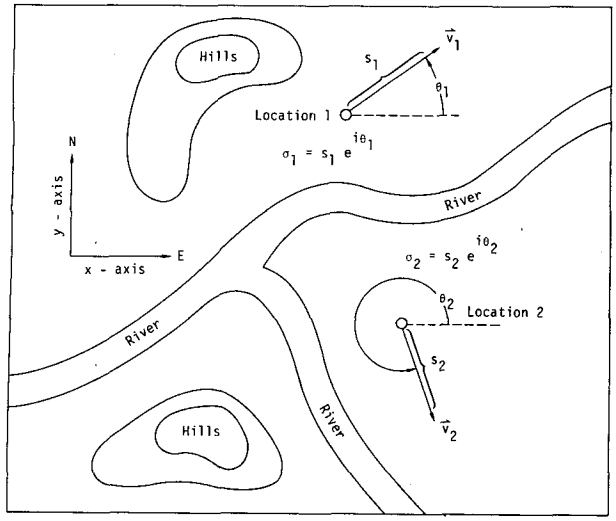

FIG. 1. A schematic representation of the relationship between wind velocities $\mathbf{v}_1$ and $\mathbf{v}_2$ observed at geographic locations 1 and 2 and the corresponding complex numbers σ_1 and σ_2 as defined by Eq. (1). Note that the arguments θ_1 and θ_2 are defined as angles measured counterclockwise from the direction of the positive x axis. The observed wind speeds at the two geographic locations are s_1 and s_2, respectively.

form the rectangular matrix. For an appropriately selected set of measurement locations and a sufficient number of measurements, the data represent a statistically significant empirical record from which the primary spatial and temporal variations of the physical parameter can be derived.

In the traditional approach, either a correlation matrix or a variance-covariance matrix is formed from the original rectangular matrix. The eigenvalues and eigenvectors of the correlation or variance-covariance matrix are then derived. The eigenvectors form a complete orthonormal set of basis vectors in terms of which the original observations can be represented. Although the significance of the eigenvalues and eigenvectors is fully covered in the references cited, their most important properties are worth enumerating here:

1) The relative magnitudes of the eigenvalues can be used to rank-order the eigenvectors in terms of decreasing significance in representing the data.

2) The most significant eigenvectors can be identified with physically important patterns in the original data.

3) Temporal expansions of the data in terms of the most significant eigenvectors are known as the principal components.

4) The primary eigenvectors provide a highly efficient basis set for approximating the original observed data.

5) The eigenvectors and principal components are determined by the original data and are derived by an objective mathematical procedure.

In the present generalization of these methods to the treatment of two-dimensional vector observations, important properties of the eigenvalues and eigenvectors as derived from scalar observations remain valid. In this paper, the required mathematical procedures are given, and the results obtained in applying the method to an extensive set of data are reported. A discussion is given in Section 4 of the application of these techniques to additional atmospheric science problems.

2. Method of analysis

This section contains a description of the mathematical procedures used. The physical significance of abstract mathematical concepts is stressed in presenting the method of analysis and in the interpretation of our results. The vector observations to be analyzed consist of wind velocities in the horizontal plane measured simultaneously at a number of geographic locations.

These data are represented as complex numbers in the analysis. Each individual observation consists of a measured wind speed and wind direction. By use of complex exponential notation, the observed two-dimensional velocity $\mathbf{v}$ is associated with a complex number

$$\sigma = se^{i\theta}, \tag{1}$$

where s is the observed wind speed and θ the observed wind direction. The meteorological convention of measuring wind directions clockwise from North is not used. The directions of the velocity $\mathbf{v}$ are defined in the present analysis as counterclockwise from the positive x axis in accordance with the usual mathematical convention. In interpreting these vectors, the positive x axis corresponds to east. The relationship between wind velocities observed at specific geographic locations and their corresponding complex numbers is indicated in Fig. 1.

Simultaneous observations at N geographic locations can be represented as a complex vector of N elements, there being a one-to-one correspondence between elements and geographic locations. The two-dimensional velocity observed at each location is associated with a complex number according to Eq. (1). All observations at time m can thus be represented as a vector $\mathbf{S}_m$ with elements σ_{km}. Each element is complex; for each time m, the magnitude is the observed wind speed and the argument the observed wind direction at the kth geographic location.

The complete set of observations is represented as an $N \times M$ rectangular matrix $\mathbf{S}$. The elements of $\mathbf{S}$ are

$$s_{km} \exp(i\theta_{km}),$$

where $m=1, M$ denotes the time of the observation, $k=1, N<M$ denotes the geographic location of the observation, s_{km} denotes the observed speed and θ_{km}

denotes the observed direction. A matrix $\mathbf{H}$ is defined in terms of $\mathbf{S}$ as

$$\mathbf{H}=M^{-1}\mathbf{S}\mathbf{S}^{\dagger}, \tag{2}$$

where $\mathbf{S}^{\dagger}$ denotes the complex conjugate transpose of $\mathbf{S}$. It follows from the definition of $\mathbf{H}$ that it is Hermitian and of dimension $N\times N$.

In the application of principal components analysis to scalar observations, the eigenvalues and eigenvectors of a correlation or a variance-covariance matrix are determined. In a similar manner, standard mathematical procedure (Smith *et al.*, 1974) can be used to determine the eigenvalues and eigenvectors of the complex matrix $\mathbf{H}$. Because $\mathbf{H}$ is Hermitian, all eigenvalues λ_j of $\mathbf{H}$ are real, and N orthonormal eigenvectors of $\mathbf{H}$ exist, satisfying

$$\mathbf{H}\hat{\mathbf{E}}_j=\lambda_j\hat{\mathbf{E}}_j, \quad j=1, \ldots, N \tag{3}$$

and the orthonormality condition

$$\hat{\mathbf{E}}_i^{\dagger}\cdot\hat{\mathbf{E}}_j=\delta_{ij}, \quad i,j=1, \ldots, N, \tag{4}$$

where $\hat{\mathbf{E}}_i^{\dagger}$ denotes the complex conjugate transpose of $\hat{\mathbf{E}}_i$ and δ_{ij} is the Kronecker delta function. Note that if θ_k is an arbitrary phase factor associated with the kth eigenvector, the set of vectors defined by

$$\hat{\mathbf{B}}_k=\exp(i\theta_k)\hat{\mathbf{E}}_k, \quad k=1, \ldots, N \tag{5}$$

are also eigenvectors of $\mathbf{H}$ and satisfy the orthonormality condition.

The eigenvectors $\hat{\mathbf{E}}_j$ are of dimension N and the elements of $\hat{\mathbf{E}}_j$ are, in general, complex. For a given eigenvector $\hat{\mathbf{E}}_j$, the kth element of $\hat{\mathbf{E}}_j$, denoted by ϵ_{kj}, is associated with the kth geographic location where speed and direction observations were recorded. Because ϵ_{kj} is complex, it may be written as

$$\epsilon_{kj}=a_{kj}\exp(i\alpha_{kj}), \quad k=1, \ldots, N, \tag{6}$$

where the magnitude a_{kj} corresponds to wind speed (expressed in relative units), and the argument α_{kj} corresponds to direction. The values of a_{kj} and α_{kj} can be used to plot N two-dimensional vectors at locations in a plane corresponding to the geographic locations at which observations were taken. Thus, each complex eigenvector $\hat{\mathbf{E}}_j$ has a geometric representation as a set of N two-dimensional vectors defined at the geographic locations where data were recorded. These eigenvectors are therefore known as spatial eigenvectors. Geometric representations of spatial eigenvectors are shown in Section 3.

The full set of N eigenvectors constitute a complete orthonormal basis set in N space. As such, they can be used to expand the data at each time m when the observations are written as an N-dimensional complex vector $\mathbf{S}_m$. It follows from the orthonormality condition (4) and the completeness of the set $\{\hat{\mathbf{E}}_j\}$ that

$$\mathbf{S}_m=\sum_{k=1}^{N} c_{km}\hat{\mathbf{E}}_k, \quad m=1, \ldots, M, \tag{7}$$

where

$$c_{km}=\hat{\mathbf{E}}_k^{\dagger}\cdot\mathbf{S}_m, \quad k=1, \ldots, N; \quad m=1, \ldots, M. \tag{8}$$

The expansion coefficients c_{km} are unique and, in general, complex.

Inspection of Eq. (7) shows that (for each time m) the magnitude of c_{km} scales each element of $\hat{\mathbf{E}}_k$ by a common multiplicative factor and that the argument of c_{km} adds a common phase angle to each element of $\hat{\mathbf{E}}_k$. In the geometric representation of an eigenvector as a set of N two-dimensional vectors, addition of a common phase angle rotates each two-dimensional vector by a common angle in the x-y plane. Thus, expansion in $\hat{\mathbf{E}}_k$ represents a scaling and a rotation of the elements of $\hat{\mathbf{E}}_k$ that is the same for all elements at each time m. The series of coefficients c_{km}, $m=1, \ldots, M$ is a time-series expansion of the observations in terms of $\hat{\mathbf{E}}_k$ and is known as the kth principal component.

Because the magnitude of c_{km} represents a relative scaling of $\hat{\mathbf{E}}_k$ when the observed data are expanded in terms of the set $\{\hat{\mathbf{E}}_j\}$, it is of interest to consider the mean-square value of c_{km}. The mean-square value of c_{km} taken over all times $m=1, \ldots, M$ gives a measure of the relative importance (scaling) of eigenvector $\hat{\mathbf{E}}_k$ in expanding the observed data. It is shown in the Appendix that

$$M^{-1}\sum_{m=1}^{M}|c_{km}|^2=\lambda_k, \quad k=1, \ldots, N. \tag{9}$$

The relative values of the eigenvalues λ_k are therefore directly related to the mean-square significance of their corresponding eigenvectors in a time-series expansion of the observed data. In principal components analysis of scalar observations, the relative values of λ_k can, in an analogous manner, be related to the reduction in variance associated with each eigenvector. A further discussion of these topics may be found in King (1969) and in Essenwanger (1976). In practical problems, the eigenvalues usually differ significantly in size. Eigenvectors corresponding to the largest eigenvalues are the primary spatial eigenvectors objectively derived from the observed data.

Associated with each spatial eigenvector is a series of expansion coefficients that represents a time-dependent modification of that spatial eigenvector. If the observed data contain recurrent temporal variations (e.g., diurnal changes), then the expansion coefficients will reflect these temporal patterns. In the present case, observations, and hence expansion coefficients, are available by hour of the day. Examples of recurrent diurnal effects as reflected in the expansion coefficients of a primary spatial eigenvector are given in Section 3.

3. Analysis and results

The techniques described in Section 2 were used to perform an objective analysis of mesoscale wind fields

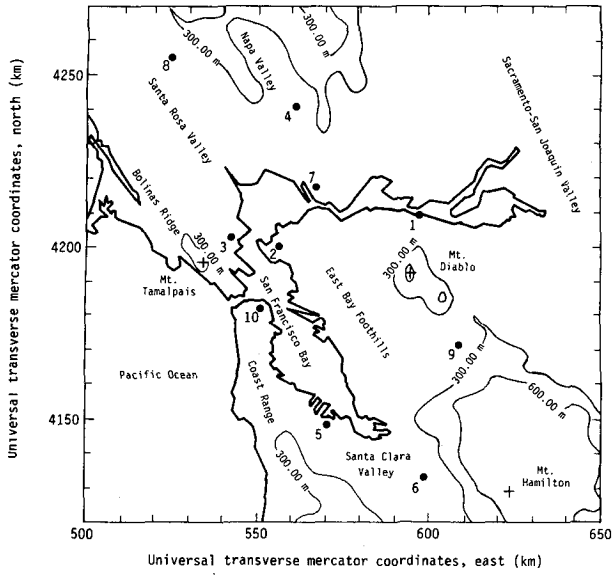

FIG. 2. The geographic locations of the ten wind velocity measurement stations that provided data for the present analysis. Topographic contours and important geographical features are identified.

in the San Francisco Bay Area, a region known for its topographic and meteorological complexity. This region includes prominent mountains, several mountain ranges, hilly areas, inland valleys, bays and rivers. Dry and wet periods typically occur each year as a result of seasonal meteorological changes. Regional winds during a given day are strongly influenced by complex topography and complicated inversion profiles over the area.

Wind velocity data for the San Francisco Bay Area were obtained from the Bay Area Air Pollution Control District (BAAPCD) network of regional measurement stations. Data selected for analysis span a 12-month period from April 1973 through March 1974. The mean hourly wind speed and direction at 10 locations were used. These data were reduced by the BAAPCD from strip-chart records and reported as mean wind speeds to the nearest mile per hour and mean wind directions by eight points of the compass. All data represent near-surface observations. Simultaneous data for all ten measurement stations were available for 7759 of the 8760 h in the 12-month period.

In Fig. 2, which shows a square portion of the Bay Area 150 km on a side, the geographic locations of the measurement stations are specified. Topographic contours indicate important terrain features. Note that the measurement stations are at low elevations, where near-surface wind velocities will often reflect the influence of valleys, mountains and passes on the regional flow.

Stations 1, 2 and 7 border areas of very low elevation along San Pablo Bay, Suisun Bay and the Sacramento and San Joaquin Rivers. Station 3 is largely blocked from west winds by Mt. Tamalpias and the Bolinas Ridge. Stations 5 and 6 are similarly blocked from west winds by the coastal and Santa Cruz Mountains. Station 9 in the Livermore valley is separated from stations 1, 2, 5 and 6 by foothills and mountains. Two stations are located in narrow valleys: Station 6 in the Santa Clara valley and station 4 in the Napa valley. Station 8 is located along the edge of the Santa Rosa valley. Mt. Diablo (located between stations 1 and 9) and Mt. Hamilton are two of the higher elevations in the area. Station 10 in the city of San Francisco is influenced by winds passing through the Golden Gate.

For convenience only, the data were organized by month and a separate analysis performed for each month of the 12-month period. The maximum size of the data matrix $\mathbf{S}$ was 10×744, corresponding to the ten measurement stations (which were numbered arbitrarily) and the maximum number of hours in a 31-day month. The $\mathbf{H}$ matrix defined by Eq. (2) for each month was of dimension 10×10. Ten complex eigenvectors, each ten-dimensional, and their associated eigenvalues were derived for each of the $\mathbf{H}$ matrices. For a given $\mathbf{H}$ matrix, the magnitudes of the eigenvalues were used to rank-order the eigenvectors in decreasing order of mean-square significance. The eigenvalues (and their associated eigenvectors) were numbered in order of decreasing magnitude such that the primary eigenvector corresponds to the largest eigenvalue λ_1, the secondary eigenvector corresponds to the next largest eigenvalue λ_2, and so forth.

For each eigenvector, a figure of merit F_k was calculated as

$$F_k=\lambda_k\Big/\sum_{j=1}^{N}\lambda_j;\quad k=1,\ldots,N. \tag{10}$$

The magnitude of F_k can range from zero to unity and gives a measure of the mean-square significance of the kth eigenvector in a time-series expansion of the data. Note that the sum of F_k over all values of k is unity. If the data were exactly represented at all hours as the product of an expansion coefficient and the primary eigenvector, the value of F_1 would be unity. The figure of merit is mathematically analogous to the fractional reduction in variance associated with each eigenvector when a variance-covariance matrix is defined from scalar data.

Table 1 lists the values of F_1, F_2 and their sum for each month. Note that F_1 is much larger than F_2 for all months, that the largest value of F_1 occurs in August and the smallest value in October and December, and that the sum of F_1 and F_2 is largest in August and smallest in December.

These results clearly reflect the basic differences between the dry period and wet period in the San Francisco Bay Area. August is included in the dry period when very regular meteorological cycles persist.

TABLE 1. The figure of merit as defined by Eq. (10) for the primary spatial eigenvector ($\mathbf{F}_1$) and the secondary spatial eigenvector ($\mathbf{F}_2$) for each month. The magnitude of the figure of merit ranges form zero to unity and gives a measure of the significance of the corresponding eigenvector as discussed in the text.

Date		Figure of merit		
Month	Year	$\mathbf{F}_1$	$\mathbf{F}_2$	$\mathbf{F}_1+\mathbf{F}_2$
April	1973	0.60	0.16	0.76
May	1973	0.63	0.13	0.76
June	1973	0.61	0.14	0.75
July	1973	0.65	0.14	0.79
August	1973	0.71	0.10	0.81
September	1973	0.62	0.13	0.75
October	1973	0.48	0.17	0.65
November	1973	0.64	0.10	0.74
December	1973	0.48	0.15	0.63
January	1974	0.53	0.11	0.64
February	1974	0.60	0.10	0.70
March	1974	0.62	0.10	0.72

December is included in the wet period when storms crossing the region greatly increase meteorological variability.

The figure of merit, eigenvalues, and the eigenvector elements for the three most significant eigenvectors derived for August 1973 are given in Table 2. Because the eigenvectors are orthonormal, the magnitudes of the elements are given in nondimensional units. The arguments of the elements are given in degrees. The magnitude and argument of each element thus define a two-dimensional vector to be associated with the numbered geographic locations shown in Fig. 2.

A time-series expansion of the data in terms of each eigenvector was performed for each month according to Eqs. (7) and (8). The magnitude and argument of the expansion coefficients c_{km} were plotted for all values of m, i.e., all hours in the month. Geometric representations of the eigenvectors were also plotted. Before plotting the expansion coefficients and geometric representations of eigenvectors, use was made of Eq. (5) by first multiplying each eigenvector $\hat{\mathbf{E}}_k$ by a phase factor $\exp(i\phi_k)$. The procedure used allows each two-dimensional vector plotted to assume a direction characteristic of the data for each measurement station.

The value of ϕ_k was chosen by defining ϕ_k arbitrarily, performing a time-series expansion according to Eq. (8), and then examining the argument of c_{km} for all values of m. In most cases, a single most probable value for the argument of c_{km} could be easily determined. A value of ϕ_k was then chosen for use in Eq. (5). This value was selected so that the mean value of the argument of c_{km} would become nearly zero when the data were expanded in terms of the eigenvectors $\hat{\mathbf{B}}_k$.

This procedure achieves a "preferred orientation" when plotting the geometric representation of an eigenvector as 10 two-dimensional vectors. When the geometric representation of an eigenvector is plotted in its preferred orientation, the two-dimensional vectors achieve a physically meaningful orientation with respect to the primary topographic features of the region. In the results shown, the preferred orientations indicate channeling effects along the river and valley areas. However, for some months (during the wet period of the year) two substantially different values for the argument of c_{km} were almost equally probable. In such cases, two preferred orientations exist due to a bimodal distribution of wind directions during the month at all stations. Two values of ϕ_k were then used in plotting geometric representations of the eigenvectors.

Geometric representations of the primary eigenvectors for six different months are shown in Figs. 3–5. Two-dimensional vectors are defined at locations corresponding to the 10 measurement stations to indicate the direction of the local air flow. The relative magnitudes of the vectors indicate the relative wind speeds at each location. Numbers assigned to each measurement location are given along with the figure of merit for the primary eigenvector.

Continuity in time is indicated by the primary eigenvectors shown in Fig. 3. April, June and August

TABLE 2. The figure of merit, eigenvalues and eigenvector elements for the first three eigenvectors derived for August 1973. The magnitudes of the elements are given on top in each case in relative nondimensional units. The arguments of the elements are given underneath in parentheses. The eigenvectors are orthonormal within rounding errors.

Eigenvector $\hat{\mathbf{E}}_1$	Figure of merit $\mathbf{F}_k$	Eigenvalue λ_k	Eigenvector elements for August 1973									
			1	2	3	4	5	6	7	8	9	10
$\hat{\mathbf{E}}_1$	0.71	69.7	0.48 (−15°)	0.37 (84°)	0.04 (180°)	0.23 (120°)	0.21 (68°)	0.12 (−49°)	0.47 (32°)	0.18 (56°)	0.31 (14°)	0.42 (32°)
$\hat{\mathbf{E}}_2$	0.10	10.3	0.52 (29°)	0.47 (−28°)	0.34 (−46°)	0.23 (−23°)	0.42 (56°)	0.19 (107°)	0.27 (−15°)	0.12 (−118°)	0.18 (114°)	0.13 (−112°)
$\hat{\mathbf{E}}_3$	0.05	5.3	0.39 (−71°)	0.58 (90°)	0.18 (160°)	0.27 (−102°)	0.31 (90°)	0.25 (118°)	0.24 (−121°)	0.15 (167°)	0.42 (145°)	0.04 (−1°)

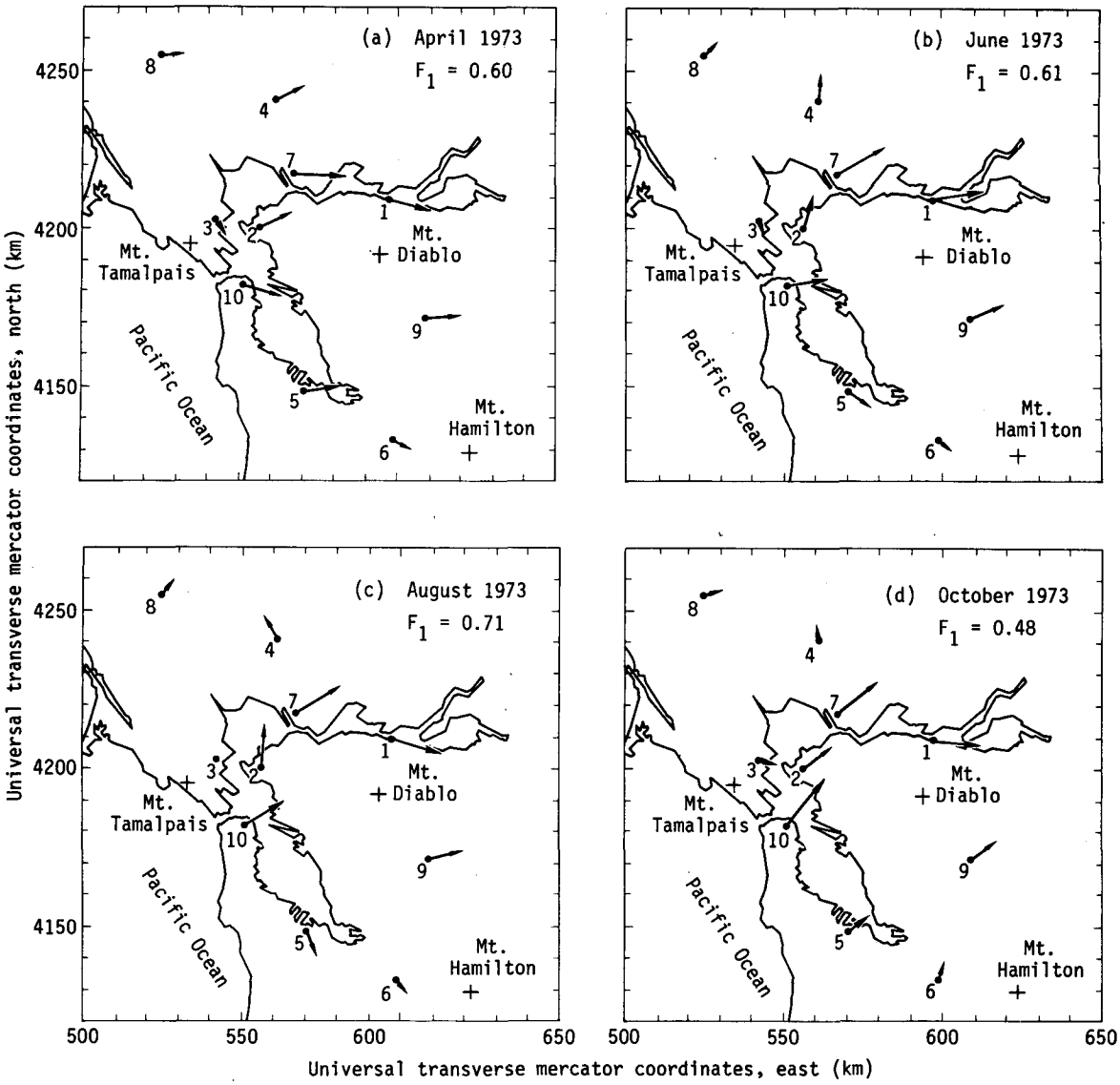

FIG. 3. Geometric representations of the primary spatial eigenvectors in their preferred orientation for April, June, August and October 1973. Two-dimensional vectors are defined at numbered locations corresponding to the ten measurement stations to indicate the direction of the local air flow. Relative wind speeds at each location are indicated by the relative magnitudes of the vectors.

1973 are very similar. Each shows strong channeling along the bay and river areas (stations 1, 2, 7). June and August show a division of the flow from station 10 around Mt. Diablo, which results in northerly winds at stations 5 and 6 and southerly winds at stations 4 and 8. October is similar to June and August, except for the emergence of a southerly flow at stations 5 and 6 and a more southerly component at stations 9 and 10.

For the months of December 1973 (Fig. 4) and February 1974 (Fig. 5), geometric representations of the primary eigenvectors are shown in two preferred orientations. For these months, two values for the argument of c_{1m} (differing by $\sim 180°$) were almost equally probable. This indicates frequent reversals of the predominant regional wind direction in which channeling effects similar to other months continue to occur. This is consistent with meteorological records for these months, which indicate the passage of storm systems across the region that resulted in regional westerly winds switching to regional easterly winds.

Regional flow reversals during the wet period of the year are common in the San Francisco Bay Area (Elford, 1974). In the winter season, the semi-permanent high-pressure area in the eastern Pacific typically moves south of its summer position, thereby permitting low-pressure storm systems to take a more southward track. Ahead of these lows, the winds are

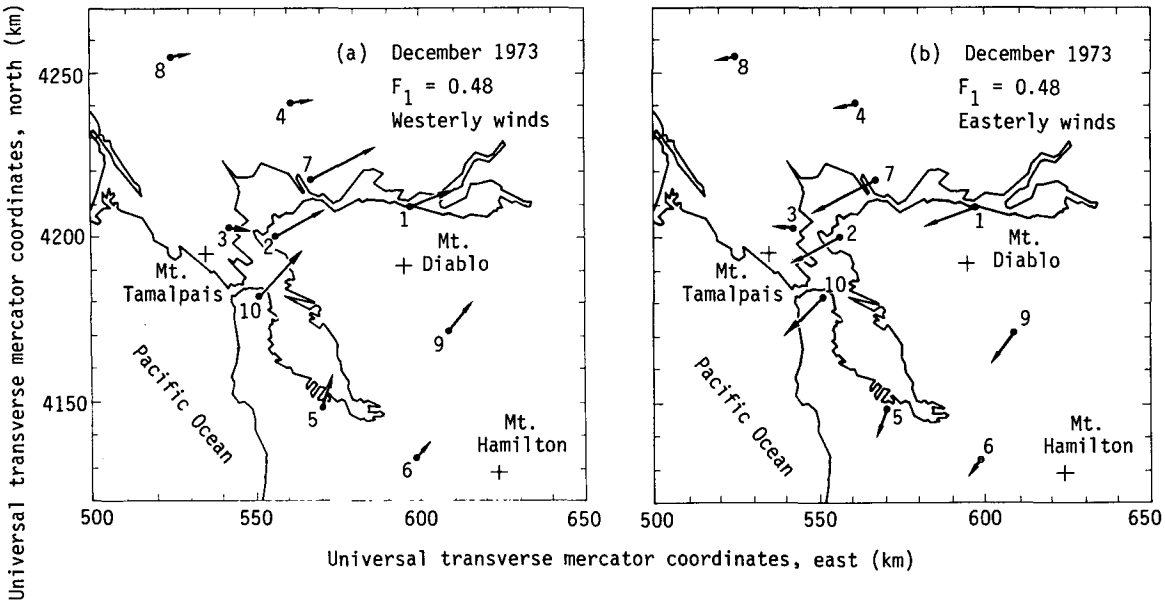

FIG. 4. Geometric representations of the primary spatial eigenvectors in two preferred orientations for December 1973. Two-dimensional vectors similar to those of Fig. 3 are defined at each measurement station.

generally westerly. Following passage of the frontal system accompanying a low-pressure area, high-pressure typically will build inland over the Great Basin. During this time, surface winds can shift through north to east in less than 24 h. Easterly winds are usually observed until the approach of another low-pressure system when winds shift back to westerly. Because of the complex topography of the San Francisco Bay region, wind shifts of up to 90° from what would be observed over uniformly flat terrain are common.

It is of interest to note that the two predominant regional patterns are very similar for December 1973 and February 1974. Derivation of the primary spatial eigenvectors for these periods makes possible a direct and quantitative comparison of the easterly and westerly patterns for these months. However, the analysis for January 1974 showed predominantly easterly regional winds without the strong tendency for regional flow reversals evident in December 1973 and February 1974. A subsequent inspection of *Daily*

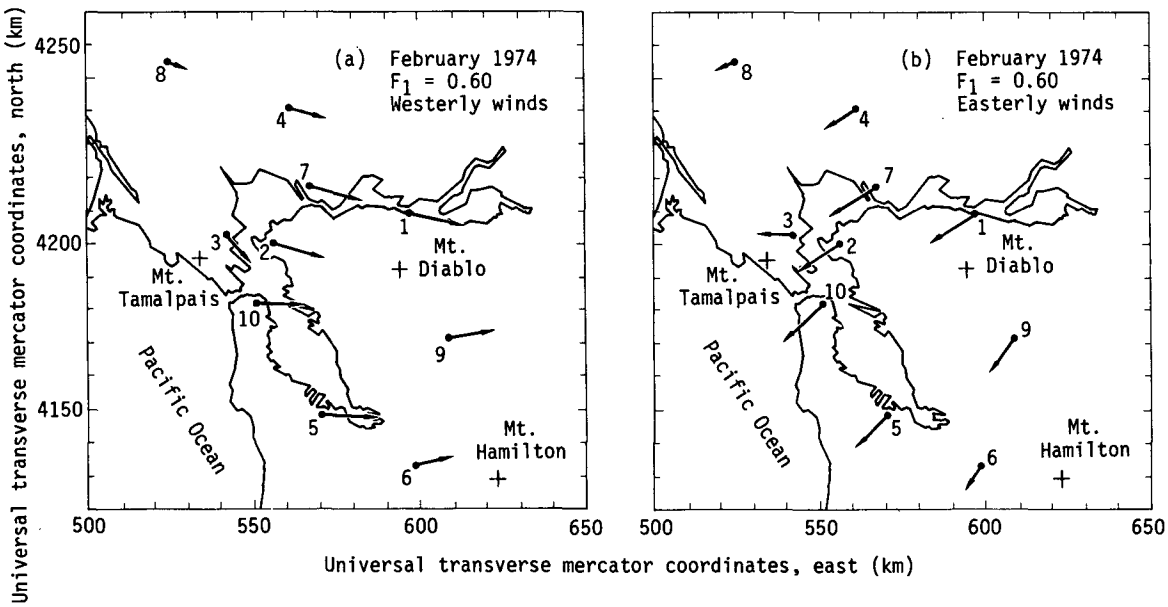

FIG. 5. Geometric representations of the primary spatial eigenvectors in two preferred orientations for February 1973. Two-dimensional vectors similar to those of Fig. 3 are defined at each measurement station.

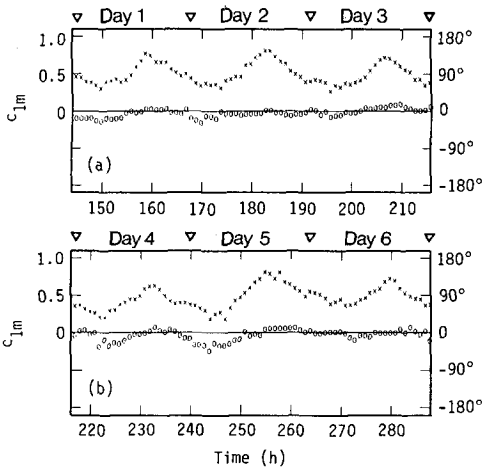

FIG. 6. The complex expansion coefficients c_{1m} corresponding to the primary spatial eigenvector $\hat{\mathbf{E}}_1$ for August 1973 vs hour of the day for two consecutive 72 h periods. The initial 72 h period is given in (a) and the following period in (b). The magnitude and argument of c_{1m} are plotted as a cross and a circle, respectively. Relative values of the magnitude are indicated along the left ordinate; values of the argument in degrees are indicated along the right ordinate. The time in hours is given on the abscissa. Triangles at the top of each figure mark the position of midnight during each day.

Weather Maps (National Oceanic and Atmospheric Administration, 1974) for January 1974 indicated an inordinate number of low-pressure systems entering the coast south of San Francisco during the first part of the month. The pressure gradient resulting from low pressure to the south was presumably responsible for a high frequency of winds directed offshore.

The primary eigenvectors represented in Figs. 3–5 indicate the primary spatial properties of the wind field. Temporal variations of the wind field are indicated by the time-dependent series of complex expansion coefficients associated with each spatial eigenvector. The series of expansion coefficients for the primary spatial eigenvector $\{c_{1m}\}$ and for the secondary spatial eigenvector $\{c_{2m}\}$ are of greatest importance.

Of the 12 months analyzed, August 1973 had the greatest diurnal regularity. Expansion coefficients associated with the primary spatial eigenvector $\hat{\mathbf{E}}_1$ for August, defined according to Eq. (8) as

$$c_{1m}=\hat{\mathbf{E}}_1^{\dagger}\cdot\mathbf{S}_m, \quad m=1, \ldots, M, \tag{11}$$

are shown in Fig. 6 for 144 consecutive hours. Fig. 6a shows an initial 72 h period and Fig. 6b the following 72 h period. The magnitude and argument of c_{1m} are separately plotted as a cross and a circle, respectively, for each hour. The position of midnight during each day is indicated. Note the regular diurnal cycle of the magnitude, which is repeated during each of the 6 days shown. This cycle represents a multiplicative scaling of the wind speed common to all 10 measurement stations. For each day, regional wind speeds reach a minimum in the early morning hours and a maximum in the mid-afternoon. Also note the relative strength of the regional wind speed for each day shown.

The argument, plotted as a circle, represents a common rotation of the two-dimensional wind velocity at all ten measurement stations. The value of the argument remains near zero because of two factors: 1) the definition of a preferred orientation for $\hat{\mathbf{E}}_1$ by means of multiplicative phase factor $\exp(i\phi_1)$ as previously discussed, and 2) the "typical" nature of the regional wind velocities observed at the ten measurement stations during the 6-day period shown. A systematic rotation of the wind direction observed at all regional measurement stations results in deviations of the argument from zero. The values of the argument are near zero during periods of maximum regional wind speeds (maximum values of the magnitude of c_{1m}). The greatest deviations of the argument from zero occur during periods of minimum wind speeds. These characteristics of the expansion coefficient series $\{c_{1m}\}$ indicate that a very regular diurnal pattern is associated with the primary spatial eigenvector for August 1973.

4. Conclusions

The results presented show that this generalization of principal components analysis can be used to advantage in studying large sets of regional wind velocity data. In particular, the method provides an objective derivation of the essential spatial and temporal properties represented in the data. This makes possible a quantitative comparison of the primary regional-scale velocity patterns on a month-by-month basis.

Regional flow reversals, such as those in December 1973 and February 1974, are also made apparent in the analysis and can be directly compared for periods of interest. In addition, regular diurnal variations of a dominant regional pattern can be objectively derived from the data.

The identification of primary spatial and temporal patterns through an objective procedure can be used to develop a quantitative prescription for "prototype days" derived from the available data. The characteristics (both in space and time) of the regional wind velocity patterns defined for such prototype days would be representative of essential quantitative features in the complete record of observations. These procedures can also be used to objectively separate data into subsets. The data in each subset might then be averaged to define typical conditions within each classification. The frequency of occurrence of classified flow types can also be easily determined.

Many applications of this analysis technique are possible. At present it is being used to classify regional wind patterns to better study spatial and temporal variations of wind energy within an area (Hardy,

1977b,c). It should be remembered, however, that the method of analysis presented is quite abstract and is not limited to studies of the type reported here. Eigenvector methods have been used to study vertical profiles of wind speed. The vector formulation of the method can now be used to analyze measured vertical profiles of wind velocity. The temporal periods treated can also be freely changed. Although periods as long as a month or more are important in developing annual assessments of either wind energy or air quality, shorter periods can be treated for other purposes. As an example, regional data for all periods preceding past air pollution episodes could be extracted from available records and analyzed. The purpose would be identification of clear regional patterns likely to signal in advance an episode of a few days duration.

Finally, it is important to note that replication of exact observations, although possible due to the completeness of the set of orthonormal basis vectors derived, is not the purpose of this analysis technique. The purpose of the analysis is to objectively derive from the original observations the essential spatial and temporal features in a quantitative form. The present generalization of principal components analysis to the treatment of two-dimensional vector observations allows an entirely new class of problems to be so treated.

Acknowledgments. We would like to recognize the valuable assistance of Kendall Peterson in providing an independent review of the San Francisco Bay Area meteorological conditions during the months treated in this study. His careful reading of the manuscript and helpful suggestions to improve and clarify the presentation of our work are also much appreciated. The contribution of the Bay Area Air Pollution Control District in providing data for analysis is also gratefully recognized. We also wish to thank Lewis Robinson of that Agency for his enthusiastic encouragement regarding development of the ideas presented here. Numerous persons, all of whom cannot be mentioned, freely gave us their help in proofreading and in editorial suggestions prior to publication. To them we are also indebted.

APPENDIX

Mathematical Relationships

The eigenvectors of the matrix $\mathbf{H}$ always form a basis set that can be used to represent the original data in accordance with Eqs. (7) and (8). An approximation to the observed data occurs when the expansion in terms of $\hat{\mathbf{E}}_k$ is truncated at $k<N$. A good approximation is obtained using only a few terms (the primary eigenvectors) whenever the eigenvalues of $\mathbf{H}$ differ significantly in size. This is true because the mean-square values of the expansion coefficients are directly related to the corresponding eigenvalues as indicated by Eq. (9). This important relationship is proved below.

The complete orthonormal basis set $\{\hat{\mathbf{E}}_k\}$ can be used to expand the data at time m such that

$$\mathbf{S}_m = c_{1m}\hat{\mathbf{E}}_1 + c_{2m}\hat{\mathbf{E}}_2 + \cdots + c_{Nm}\hat{\mathbf{E}}_N, \tag{A1}$$

where the coefficients c_{km} are given by the inner products

$$c_{km} = \hat{\mathbf{E}}_k^\dagger \cdot \mathbf{S}_m, \quad k = 1, \ldots, N. \tag{A2}$$

The absolute square of c_{km} is

$$|c_{km}|^2 = (\hat{\mathbf{E}}_k^\dagger \cdot \mathbf{S}_m)(\hat{\mathbf{E}}_k^\dagger \cdot \mathbf{S}_m)^*. \tag{A3}$$

However, for any pair of N-dimensional complex vectors $\mathbf{A}$ and $\mathbf{B}$, it follows that

$$\mathbf{A}^\dagger \cdot \mathbf{B} = \sum_{n=1}^{N} a_n^* b_n = (\mathbf{B}^\dagger \cdot \mathbf{A})^*, \tag{A4}$$

where a_n and b_n are the elements of $\mathbf{A}$ and $\mathbf{B}$, respectively. Therefore, (A3) can be written as

$$|c_{km}|^2 = (\hat{\mathbf{E}}_k^\dagger \cdot \mathbf{S}_m)(\mathbf{S}_m^\dagger \cdot \hat{\mathbf{E}}_k) \tag{A5}$$

$$= \hat{\mathbf{E}}_k^\dagger [\mathbf{S}_m \mathbf{S}_m^\dagger] \hat{\mathbf{E}}_k, \tag{A6}$$

where $[\mathbf{S}_m \mathbf{S}_m^\dagger]$ is a dyad. The elements of the dyad, given in terms of the elements of $\mathbf{S}_m$, are

$$\begin{matrix} \sigma_{1m} \\ \sigma_{2m} \\ \vdots \\ \sigma_{Nm} \end{matrix} \begin{bmatrix} \sigma_{1m}^* & \sigma_{2m}^* & \cdots & \sigma_{Nm}^* \end{bmatrix} = \begin{bmatrix} \sigma_{1m}\sigma_{1m}^* & \sigma_{1m}\sigma_{2m}^* & \cdots & \sigma_{1m}\sigma_{Nm}^* \\ \sigma_{2m}\sigma_{1m}^* & \cdots & & \\ \vdots & & & \\ \sigma_{Nm}\sigma_{1m}^* & & & \sigma_{Nm}\sigma_{Nm}^* \end{bmatrix}. \tag{A7}$$

The mean-square value of c_{km} taken over m is, by use of (A6),

$$M^{-1} \sum_{m=1}^{M} |c_{km}|^2 = \hat{\mathbf{E}}_k^\dagger \{ M^{-1} \sum_{m=1}^{M} [\mathbf{S}_m S_m^\dagger] \} \hat{\mathbf{E}}_k. \tag{A8}$$

Note that the expression within the outer brackets of (A8) is a dyadic. Individual terms in the dyadic sum over m are of the form given by (A7). By comparison with the definition of the matrix $\mathbf{S}$, it can be seen that the dyadic in (A8) is equal to $\mathbf{H}$, i.e.,

$$M^{-1} \sum_{m=1}^{M} [\mathbf{S}_m \mathbf{S}_m^\dagger] = M^{-1} \mathbf{S}\mathbf{S}^\dagger = \mathbf{H}. \tag{A9}$$

Therefore,

$$M^{-1} \sum_{m=1}^{M} |c_{km}|^2 = \hat{\mathbf{E}}_k^\dagger \mathbf{H} \hat{\mathbf{E}}_k$$

$$= \lambda_k \hat{\mathbf{E}}_k^\dagger \cdot \hat{\mathbf{E}}_k$$

$$= \lambda_k, \quad k = 1, \ldots, N, \tag{A10}$$

which completes the proof that the mean-square values of the expansion coefficients are equal to the corresponding eigenvalues.

It is also of interest to note that in the present case,

the absolute square of $\mathbf{S}_m$ is the wind speed squared summed over all geographic locations at time period m. It is given by

$$|\mathbf{S}_m|^2 = \sum_{k=1}^{N} \sigma_{km}\sigma_{km}^* = \sum_{k=1}^{N} s_{km}^2, \quad \text{(A11)}$$

where s_{km} has been previously defined. Due to orthonormality of the set $\{\hat{\mathbf{E}}_k\}$, it is seen from (A1) that

$$|\mathbf{S}_m|^2 = \mathbf{S}_m^\dagger \cdot \mathbf{S}_m = \sum_{k=1}^{N} |c_{km}|^2. \quad \text{(A12)}$$

Thus, if the mean of $|\mathbf{S}_m|^2$ overall values of m is denoted by T, then

$$T = M^{-1} \sum_{m=1}^{M} |\mathbf{S}_m|^2 \quad \text{(A13)}$$

$$= \sum_{k=1}^{N} \{M^{-1} \sum_{m=1}^{M} |c_{km}|^2\} \quad \text{(A14)}$$

$$= \sum_{k=1}^{N} \lambda_k. \quad \text{(A15)}$$

The wind speed squared summed over all geographic locations is, neglecting density variations among locations, proportional to the kinetic energy of the wind summed over the finite number of observing stations. The sum of the eigenvalues of $\mathbf{H}$ is equal to the time-average over m of this quantity.

REFERENCES

Blifford, I. H., and G. O. Meeker, 1967: A factor analysis model of large-scale pollution. *Atmos. Environ.*, **1**, 147–157.

Clark, J. F., and J. T. Peterson, 1973: An empirical model using eigenvectors to calculate temporal and spatial variations of the St. Louis heat island. *J. Appl. Meteor.*, **12**, 195–210.

Elford, C. R., 1974: Climate of California. *Climates of the States*, Vol. 2, *Western States*, Water Information Center, Inc., Port Washington, NY, 538–546.

Essenwanger, O. M., 1976: *Applied Statistics in Atmospheric Science*, Part A, Frequencies and Curve Fitting. Elsevier, 412 pp.

Hardy, D. M., 1976: Wind power studies: Initial regional applications. Lawrence Livermore Laboratory Rep. UCRL-50034-76-2.

——, 1977a: Empirical eigenvector analysis of vector observations. *Geophys. Res. Lett.*, 4, 319–320.

——, 1977b: Wind studies in complex terrain. Lawrence Livermore Laboratory Rep. UCRL-79430.

——, 1977c: Numerical and measurement methods of wind energy assessment. *Proc. Third Biennial Conf. Wind Energy Conversion Systems*, U.S. Dept. of Energy, 664–676.

Kidson, J. W., 1975: Eigenvector analysis of monthly mean surface data. *Mon. Wea. Rev.*, **103**, 177–186.

King, L. J., 1969: *Statistical Analysis in Geography*. Prentice-Hall, Chap. 7.

National Oceanic and Atmospheric Administration, 1974: *Daily Weather Maps*. U.S. Department of Commerce, Washington, DC.

Peterson, J. T., 1970: Distribution of sulfur dioxide over metropolitan St. Louis, as described by empirical eigenvectors, and its relation to meteorological parameters. *Atmos. Environ.*, 4, 501–518.

Smith, B. T., J. M. Boyle, B. S. Garbow, Y. Ikebe, V. C. Klema and C. B. Moler, 1974: Matrix eigenvalue routines—EISPACK guide. *Lecture Notes in Computer Science*, Vol. 6, Springer-Verlag.

Stidd, C. K., 1967: The use of eigenvectors for climatic estimates. *J. Appl. Meteor.*, **6**, 255–264.