An Index of Arizona Summer Rainfall Developed through Eigenvector Analysis

Daniel Morgan Johnson
Portland State University, Portland, OR 97207

(Manuscript received 1 November 1979, in final form 15 January 1980)

ABSTRACT

Summer rainfall in Arizona exhibits tremendous spatial variability due to topographic irregularities and the fact that convection is the dominant precipitation generating mechanism. As such it has proven difficult to document changes over time, changes in either total amounts or in spatial distribution. In this study, eigenvector analysis is utilized to generate an effective index of this climatic parameter. From an original data matrix consisting of summer rainfall totals at 101 climatological stations for 50 years, the first time-series eigenvector is extracted and accounts for 70% of the variance. Geographic mapping of the multipliers (coefficients) reveals that the first eigenvector does adhere to physical reality, reflecting the dominant pattern of rainfall over the state. Thus, its validity as an effective index is established.

1. Introduction

The use of eigenvectors to describe patterns or arrays of data has been extensive in meteorological and climatological research since Lorenz (1956) first used them for weather prediction. For example, Grimmer (1963) used them to explore patterns of temperature in Europe, Sellers (1968) to determine the dominant precipitation anomaly patterns for the western United States for each month over a 36-year period, and Stidd (1967) to describe the dominant patterns in the annual precipitation regime of Nevada. Stidd, by identifying three "dominant" distributional patterns which accounted for 93% of the "variance" in the original data matrix, was able to develop a very accurate technique for estimating the mean monthly precipitation for any point or area in Nevada.

More recent climatological applications of the technique are Wright's (1974) clarification of the dominant patterns of seasonal rainfall in southwestern Australia and Kalnicky's (1974) documentation of a major shift in the circulation of the northern hemisphere around 1950, a change from predominantly zonal flow to predominantly meridional. Spatial and temporal patterns of drought in the United States (Skaggs, 1975) and in the upper midwest (Klugman, 1978) were identified using monthly values of the Palmer Drought Index as the basic data.

One reason that eigenvectors have not been even more widely used is that the mathematics are not easily explained without the use of matrix theory. However, their popularity has developed because of their facility in reducing a massive array of data and because of the distinct possibility of assigning physical significance to them. Their function is to specify the timewise variation of a space field by means of a much smaller number of variables than the original number of variables; thus, they can be determined for either the time dimension or the space dimension. The purpose of this paper is to present another climatological application of eigenvector analysis; specifically, the establishment of a single index to characterize the statewide pattern of Arizona summer rainfall.

2. Arizona summer rainfall and climatic change

Arizona and the surrounding geographical area experience a distinctive bimodal distribution of annual precipitation, winter storms being dependent on the midlatitude westerly circulation. The summer regime (June–September) is more nearly a response to subtropical atmospheric controls. Winter precipitation in Arizona is the most significant factor in the availability of water in much of the state and as such has received the most attention from researchers interested in climatic change. Yet in the southeast portion of the state summer rainfall contributes more to annual precipitation totals than does winter precipitation (Fig. 1). In order to fully appreciate the role of climatic change in the state's water supply, the summer season requires investigation separate from winter.

A causal mechanism necessarily exists between the Arizona summer precipitation regime and synoptic-scale circulation patterns. It follows logically that any significant changes detected in summer rainfall, changes in the amount of rainfall and in its spatial and temporal occurrence, would imply

0021-8952/80/070849-08$06.00

important fluctuations in controlling mechanisms. However, the problem of quantifying the spatial distribution and changes over time is made exceedingly difficult by the great geographical variability in this parameter (Fig. 2); thus the application of eigenvector analysis to establish a statistical index which adheres to physical reality.

3. Data

All available climatological stations in Arizona and adjacent areas of surrounding states were inventoried to determine the length and quality of summer precipitation records. It was found that the adoption of a 50-year base period (1927–76) would allow maximum use of the data. The validity of the statistical characteristics of any climatic parameter increases correspondingly with the length of the record. However, the use of a base period exceeding 50 years would cause a rapid reduction in the number of continuous station records available. Eighty-two stations in Arizona were thus selected, supplemented with 19 from the surrounding states, to yield a total network of 101 stations, displayed in Fig. 3. The basic data, then, is the time series of summer rainfall totals comprised of the sum of monthly values from June–September.

4. Analysis

The eigenvectors of a variance-covariance matrix are analogous to the principal components of a data

FIG. 1. Arizona summer precipitation (June–September) expressed as a percentage of annual precipitation, 1927–76.

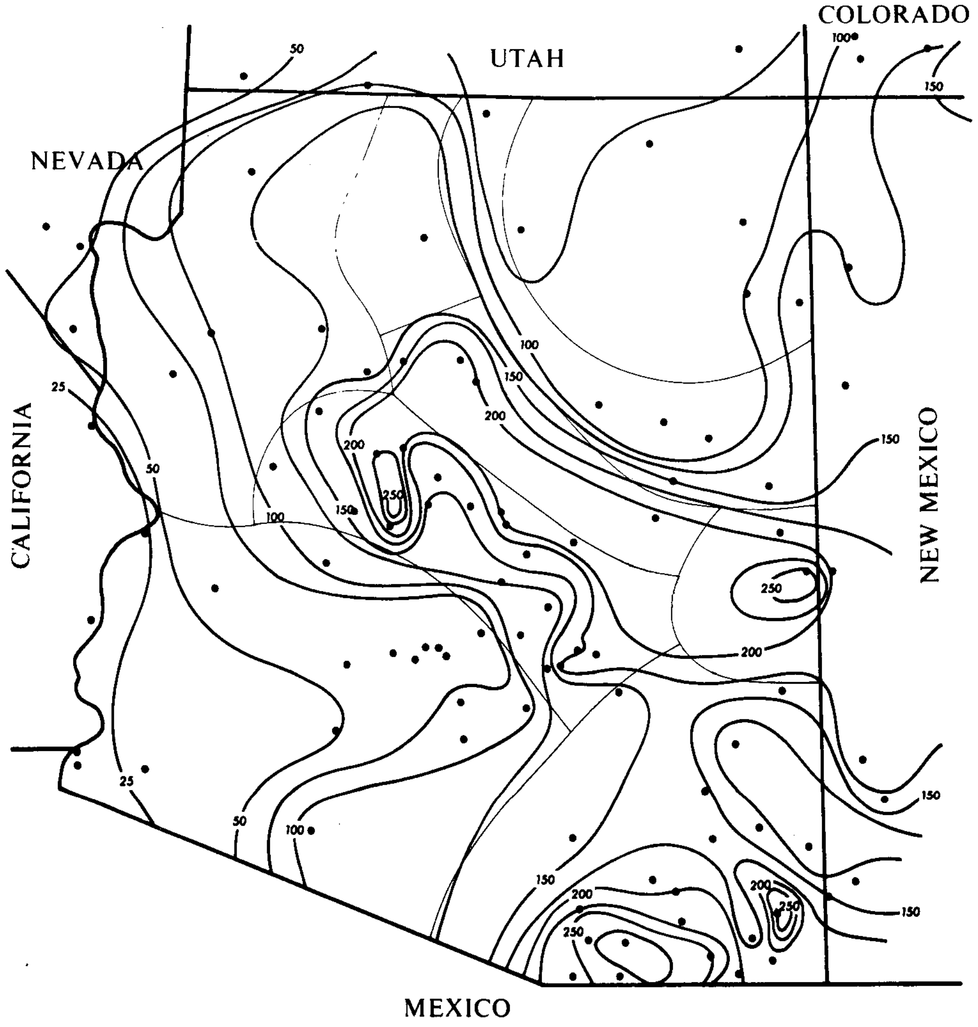

FIG. 2. Median values of summer rainfall (depth in mm) for Arizona and adjacent regions, 1927–76.

set. The data matrix **A** in this study consists of the summer rainfall totals at 101 stations over a 50-year period (1927–76); i.e., **A** is a 50 (row) × 101 (column) matrix, where one of the elements a_{ij} is the summer season precipitation for the ith year at the jth station. A simpler way of visualizing this matrix is to consider each column as a time series of 50 years at that particular station. Multiplication of the matrix **A** by its transpose $\mathbf{A}^T$ yields a new matrix **B** which is the variance-covariance matrix of the original data set. Matrix **B**, therefore, is a real, symmetric matrix of 50 rows and 50 columns.

A visual interpretation is useful at this point for the interested reader. Elements in an $m \times m$ matrix can be regarded as defining points lying on an m-dimensional ellipsoid. This representation can be visualized for two and possibly three dimensions; but it is impossible for a larger matrix and can only be considered mathematically. The variance-covariance matrix **B** of 50 dimensions is of this type. The eigenvectors of the matrix define the principal axes of the multi-dimensional ellipsoid, and the eigenvalues represent the lengths of these axes.

Computation of the eigenvectors and eigenvalues for a large matrix necessitates the use of a high-speed computer (Dixon, 1973). Briefly, from the variance-covariance matrix **B** a determinant is formed. Solution of $\det(\mathbf{B} - \lambda\mathbf{I}) = 0$ yields the eigenvalues of the matrix which are the elements of λ, otherwise known as the scalar characteristic roots. **I** is the identity matrix. Because a variance-covariance matrix is always symmetrical, these 50 eigenvectors will be orthogonal; i.e., oriented at right angles to each other.

The eigenvectors, then, are a new set of axes which represent the original data. In a 50-dimensional space, the 101 climatological stations may theoretically be located by their respective summer

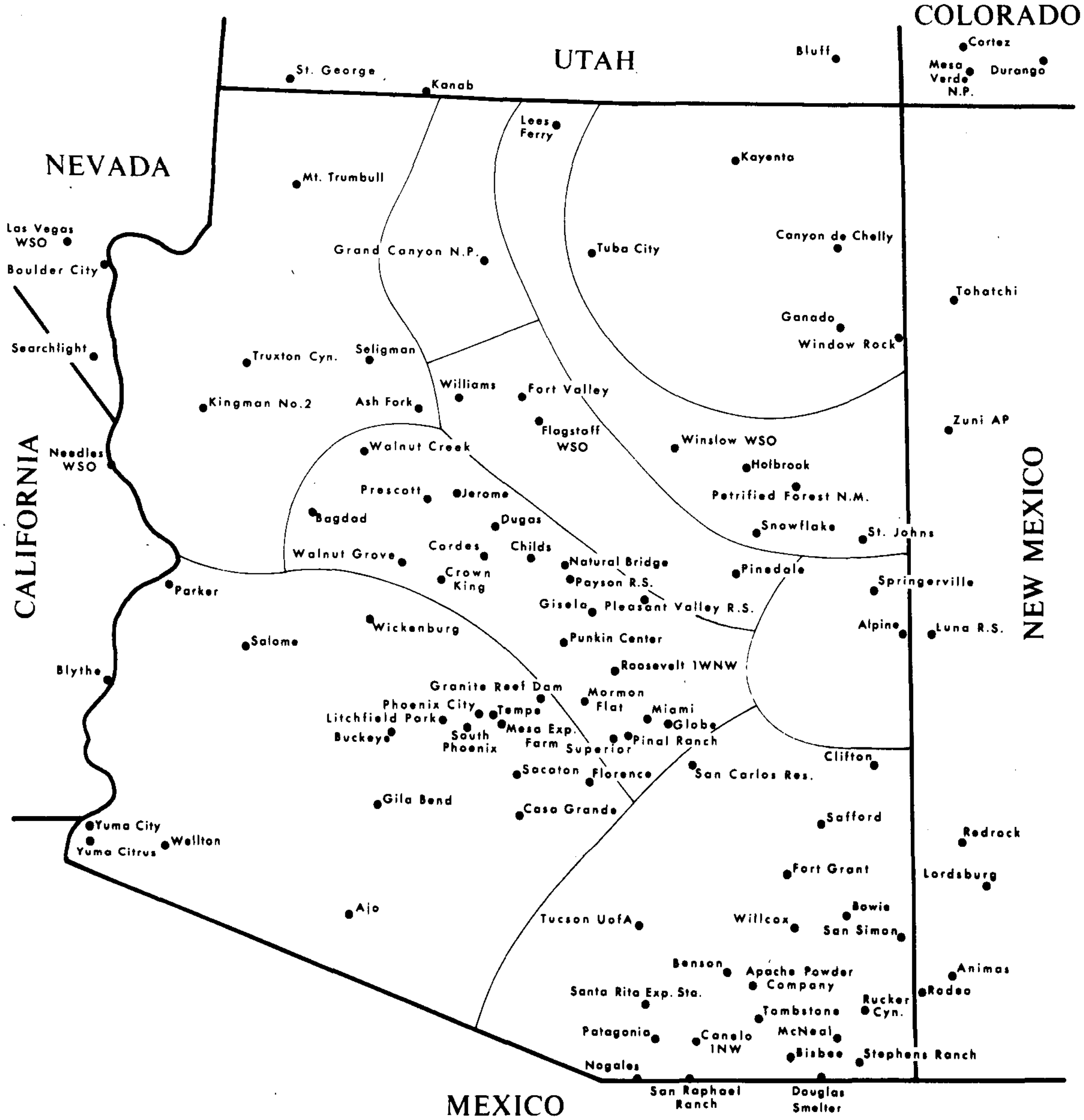

FIG. 3. Map of Arizona and adjacent regions showing the location of the 101 climatological stations used as a data base in this study.

TABLE 1. The set of the first ten eigenvalues extracted from Arizona summer rainfall totals.

Eigenvalue	Magnitude	Cumulative proportion of total variance
1	381.291	0.696
2	23.557	0.739
3	14.341	0.766
4	10.506	0.785
5	8.673	0.801
6	7.302	0.814
7	6.919	0.827
8	6.236	0.838
9	5.737	0.848
10	5.593	0.859

rainfall totals for each of the 50 years. Stations with similar histories of wetness and dryness will be close together in this hyperspace. Certainly, many different combinations may exist. Computed eigenvalues extracted from matrix **B** are arranged in order of magnitude in Table 1. Also listed is the cumulative proportion of the total variance accounted for. The eigenvector associated with the largest eigenvalue is the vector from the origin through the point in hyperspace which maximizes the variance explained. The eigenvector associated with the second largest eigenvalue maximizes the remaining variance to be explained and is orthogonal to the first eigenvector. The remaining eigenvectors have the same characteristics. If all the eigenvalues were listed in Table 1, instead of only the first ten, 100% of the variance would be accounted for.

It can be seen that the eigenvector associated with the first eigenvalue accounts for nearly 70% of the variance in the original data matrix. Although large, this value is not unexpected considering the source of the data. Summer rainfall over the state reflects primarily one dominant control—the influx of tropical moisture from the Gulf of Mexico (Bryson and Lowry, 1955) and the North Pacific via the Gulf of California (Hales, 1972, 1974) due to the relative location and intensity of the two subtropical anticyclones. The first eigenvalue/eigenvector represents the pattern of variation due to this dominant influence while the others are progressively local patterns, explain only small percentages of the remaining variance, and are nearly indistinguishable from each other in this regard. Thus, only the eigenvector associated with this first eigenvalue was retained for further analysis, being the only one with any potential as an index of the amount and spatial distribution of summer rainfall.

Additional information is obtained by using $\mathbf{E}^T \times \mathbf{A} = \mathbf{M}$ where $\mathbf{E}^T$ is the transpose of the eigenvector matrix. If only the first eigenvector is considered, then $\mathbf{E}^T$ is a (1×50) matrix, $\mathbf{A}$ being the original data matrix (50×101). $\mathbf{M}$ is a (1×101) matrix of coefficients, or multipliers as Stidd (1967) chooses to call them. Each eigenvector for each data point (climatological station) has one unique multiplier. In reference to the aforementioned visual description, this procedure projects each observation onto the principal axis (first eigenvector). Multipliers, then, are simply the coefficients of the linear equation which the eigenvector defines. In a more applied sense, they represent the degree of eigenvector “importance” at each data point. The 101 multipliers of the first eigenvector have been plotted on a map at the corresponding climatological sta-

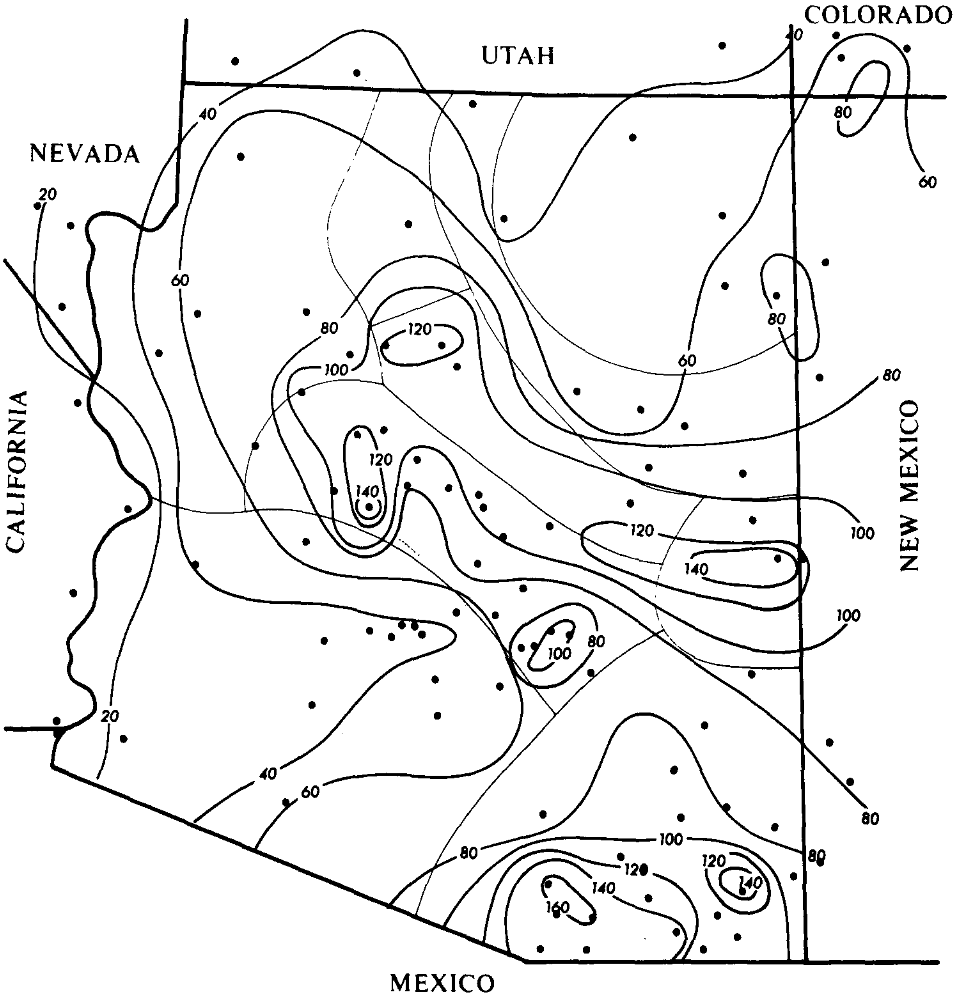

FIG. 4. Geographical distribution of the multipliers of the first eigenvector, 1927–76. Magnitude indicates relative degree of importance.

tions as shown in Fig. 4. It is encouraging to see the same pattern emerge as with the map of median seasonal rainfall for the same 50-year base period, evidence that this pattern of multipliers does, indeed, represent the dominant pattern of distribution.

However, the selected stations vary considerably in elevation, from near sea level to over 2440 m. Since, in Arizona, summer rainfall is strongly correlated with elevation, it was logical to assume that the multipliers themselves might be correlated with elevation. Simple linear regression was used to define this relationship and to remove at least a part of the elevation effect. The correlation coefficient (r) with elevation = 0.62. Correction of all multipliers to a common elevation of 1220 m MSL resulted in a slightly different spatial pattern, a more meaningful pattern of eigenvector "importance" as displayed in Fig. 5.

In contrast to Fig. 4, a southeast maximum now becomes more outstanding. The core is an area centered around Nogales, on the Arizona-New Mexico border, where values are in excess of 140. The analysis thus reflects what is commonly known regarding the preferred route of tropical moisture into the state. As described by Bryson and Lowry (1955) the onset of significant amounts of summer rainfall in early July is the result of a rather sharp transition from one dominant air mass to another over the state. The transition is coincident with a major readjustment in the general circulation including a rather sudden shift in the mean latitude of the center of the eastern Pacific anticyclone from about latitude 34° north to about 40° north. Simultaneously, the Bermuda high in the North Atlantic expands, intensifies, and moves northward, causing moist air to move into the state on a broad band of southeast

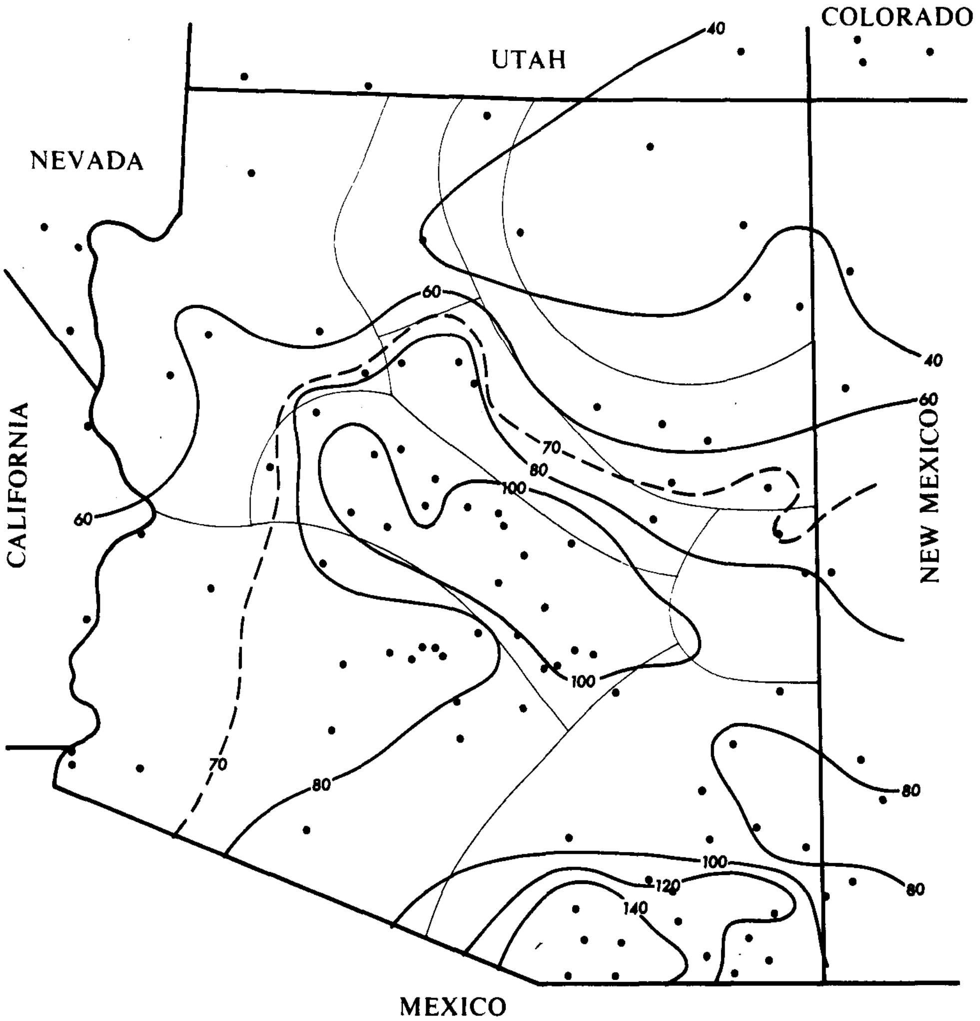

FIG. 5. Geographical distribution of the multipliers of the first eigenvector, corrected to a common elevation of 1220 m MSL.

winds, moving anticyclonically over the Mexican highlands and into the intermountain regions of the southwest (see Fig. 6). Initial entry into the state is generally in that area highlighted by maximum **M** values in Fig. 5. The rainy season terminates more gradually in September as extratropical westerly flow becomes increasingly present over the state. The end of the summer rainy reason is frequently punctuated by heavy rainfall events in the southern half of the state due to decaying tropical storms from the eastern North Pacific.

To present an example of the physical validity of this pattern of multipliers, it is noted that the southwestern quarter of the state, which receives minimal amounts of rainfall, displays higher M-values than does the northern third of the state. The more northerly regions are further from the principal moisture source and receive a proportionately greater share of their moisture from other sources such as midlatitude westerly flow in early and late summer. Thus a significant physical interpretation of the technique is that the relative effectiveness of the dominant influence, the influx of tropical moisture into the state, is rapidly lowered with northward progression. The validity of the first time-series eigenvector is thus established and it can be used as an effective index of ''summer'' rainfall over the entire state.

5. Discussion

Most investigations of recent climatic change are based on detailed analysis of data for relatively few stations which are considered representative of the regions within which they occur. Few studies have been concerned with detailed regional climatic change based on the analysis of data for a dense network of stations (Tyson *et al.*, 1975). This analysis has demonstrated that the first eigenvector obtained from principal components analysis clearly delineates those parts of Arizona most affected by the dominant precipitation pattern for the summer period. As such this index of the dominant spatial pattern has utility in the study of climatic change. An eigenvector, as calculated in this study, consists of 50 dimensions and can be interpreted as a 50-term time series of a spatially weighted mean for the entire region (see Fig. 7). High values imply that wetter than normal conditions prevailed, while low values indicate drier years, a relationship verified by the mapping of statewide rainfall for each of the 50 years, 1927–76.

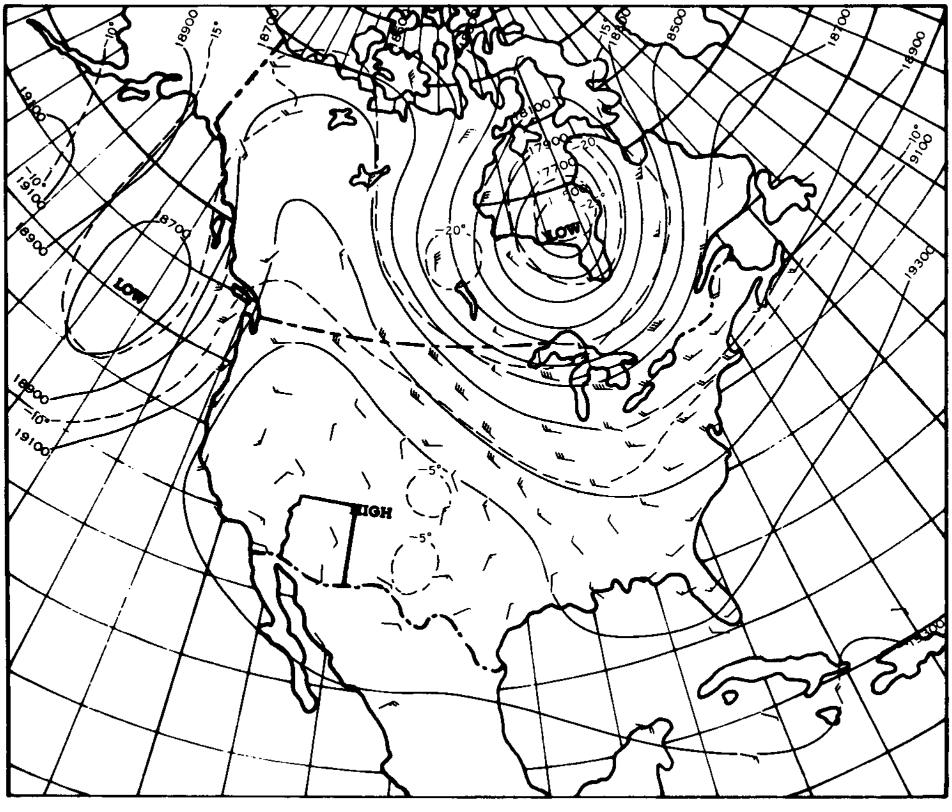

FIG. 6. Airflow into Arizona and the Southwest during the summer rainfall season as illustrated by the synoptic map of 500 mb height contours, 9 July 1975 (National Oceanic and Atmospheric Administration, 1975).

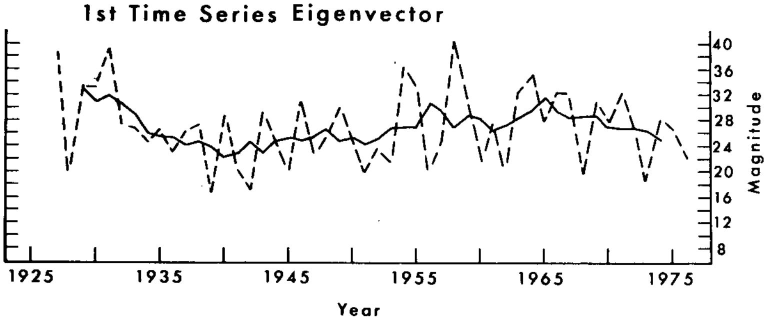

FIG. 7. Time series plot of the first eigenvector of Arizona summer rainfall, 1927–76. High values correspond with wet years; low values correspond with dry years. Solid line is a 5-year running mean.

Analysis of the data in Fig. 7 is beyond the intent of this paper. The point to be made here is that, given the high degree of spatial variability in Arizona summer rainfall, it is impossible to generate a reliable time-series of regional rainfall by the traditional technique of averaging several stations. The first eigenvector accomplishes this task with 70% efficiency.

Time series analysis applied to the index would yield more meaningful results than the traditional analysis of individual stations. The disadvantage, of course, is that while individual stations in Arizona may approach, and in some cases exceed, 100 years of record, the eigenvector index, requiring a dense network of stations, is limited to only the more recent part of the climatological record. Certainly, the potential for use of eigenvectors as a tool in the study of climatic change is great. However, in each case their spatial validity and physical reality must first be established.

Acknowledgments. This research was accomplished at the Laboratory of Climatology, Arizona State University, under the general guidance of Drs. Robert W. Durrenberger and Anthony J. Brazel. Special thanks is given to Dr. George Hepner, Western Michigan University, for his advice and counsel regarding the statistical analysis. Financial support for publication costs was generously provided by the Portland State University Foundation.

REFERENCES

Bryson, R. A., and W. P. Lowry, 1955: The synoptic climatology of the Arizona summer precipitation singularity. *Bull. Amer. Meteor. Soc.*, **36**, 329–339.

Dixon, W. J., 1973: *BMD; Biomedical Computer Programs*, 3rd ed. Berkeley, University of California Press, 585 pp.

Grimmer, M., 1963: The space-filtering of monthly surface anomaly data in terms of pattern, using empirical orthogonal functions. *Quart. J. Roy. Meteor. Soc.*, **89**, 395–408.

Hales, J. E., 1972: Surges of maritime tropical air northward over the Gulf of California. *Mon. Wea. Rev.*, **100**, 298–306.

——, 1974: Southwestern United States summer monsoon source—Gulf of Mexico or Pacific Ocean? *J. Appl. Meteor.*, **13**, 331–342.

Kalnicky, R. A., 1974: Climatic change since 1950. *Ann. Amer. Assoc. Geogr.*, **64**, 100–112.

Klugman, M. R., 1978: Drought in the upper midwest, 1931–1969. *J. Appl. Meteor.*, **17**, 1425–1431.

Lorenz, E. N., 1956: Prospects for statistical weather forecasting. Final Report, Statistical forecasting project, Massachusetts Institute of Technology, 103 pp.

National Oceanic and Atmospheric Administration (NOAA), 1975: *Daily Weather Maps*. Washington DC, Govt. Printing Office.

Sellers, W. D., 1968: Climatology of monthly precipitation patterns in the western United States, 1931–1966. *Mon. Wea. Rev.*, **96**, 585–595.

Skaggs, R. H., 1975: Drought in the United States. *Ann. Amer. Assoc. Geogr.*, **65**, 391–402.

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

Tyson, P. D., T. G. J. Dyer and M. N. Mametse, 1975: Secular changes in South African rainfall: 1880 to 1972. *Quart. J. Roy. Meteor. Soc.*, **101**, 817–833.

Wright, P. B., 1974: Seasonal rainfall in southwestern Australia and the general circulation. *Mon. Wea. Rev.*, **102**, 219–232.