

AD-A047 791

AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OHIO
AN EXPLORATORY ANALYSIS OF AN EIGENVECTOR-BASED RECONSTRUCTION --ETC(U)
JUN 76 E W HOLLINGSHEAD
AFIT-CI-78-13

F/6 4/2

UNCLASSIFIED

NL

| OF |
AD
A047791



END
DATE
FILMED
1 -78
DDC

AD A 0 477 91

①
B. S.

78-13

**AN EXPLORATORY ANALYSIS OF AN EIGENVECTOR-
BASED RECONSTRUCTION MODEL FOR THE
SOUTHWEST MONSOON PRECIPITATION RECORDS
OF THE INDIAN SUBCONTINENT**

by

Edward William Hollingshead

DDC
RECEIVED
DEC 21 1977
AT E

**DEPARTMENT OF METEOROLOGY
University of Wisconsin—Madison**

1976

AD No. _____
DDC FILE COPY

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER AFIT - CI-78-13		2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) An Exploratory Analysis of an Eigenvector-Based Reconstruction Model for the Southwest Monsoon Precipitation Records of the Indian Subcontinent.		5. TYPE OF REPORT & PERIOD COVERED Master's Thesis	
7. AUTHOR(s) 10. William Edward W. Hollingshead		6. PERFORMING ORG. REPORT NUMBER	
9. PERFORMING ORGANIZATION NAME AND ADDRESS AFIT Student at the University of Wisconsin, Madison, Wisconsin		8. CONTRACT OR GRANT NUMBER(s)	
11. CONTROLLING OFFICE NAME AND ADDRESS AFIT/CI WPAFB OH 45433		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS	
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE 11. Jun 76	
		13. NUMBER OF PAGES 75 Pages 10. 85 p.	
		15. SECURITY CLASS. (of this report) Unclassified	
16. DISTRIBUTION STATEMENT (of this Report) Approved for Public Release; Distribution Unlimited		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE	
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		DDC RECEIVED DEC 21 1977 RECEIVED F	
18. SUPPLEMENTARY NOTES APPROVED FOR PUBLIC RELEASE AFR 190-17. JERRAL F. GUESS, Captain, USAF Director of Information, AFIT		Ø 12 200	
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)			
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)			

— (1)

**An Exploratory Analysis of an Eigenvector-Based Reconstruction Model
for the Southwest Monsoon Precipitation Records
of the Indian Subcontinent**

by

Edward William Hollingshead

A thesis submitted in partial fulfillment of the requirements
for the degree of

**Master of Science
(Meteorology)**

at the

University of Wisconsin-Madison

1976

DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited

AN EXPLORATORY ANALYSIS OF AN EIGENVECTOR-BASED RECONSTRUCTION MODEL FOR
THE SOUTHWEST MONSOON PRECIPITATION RECORDS OF THE INDIAN SUBCONTINENT

Edward William Hollingshead

Under the supervision of Professor Eberhard W. Wahl

↘ A 53 station southwest monsoon precipitation record for all Julys, 1921-1960, is cube rooted, normalized and subjected to an eigenvector analysis to resolve the spatial pattern of precipitation. The first four eigenvectors, accounting for 44% of the variance, best portray the macroscale pattern of the July monsoon with an apparent minimum amount of interference from local effects. ↗

A method of reconstructing an approximate precipitation record using either four or thirteen eigenvectors for the orthogonal base of a regression scheme is discussed. The 1921-1960 eigenvectors appear to be acceptable as the orthonormal base for years outside the data set. Use of all thirteen eigenvectors provides the better estimate of the departure pattern when only a few stations are missing from the data set, while use of the first four eigenvectors produces better results when a significant fraction of the stations are missing from the data set. The four eigenvector model (and to a lesser extent the thirteen eigenvector model) generates estimates that are nearly as accurate for the stations excluded from the regression as for those included in the regression equation. ↗

Approved by

Edward W. Hollingshead

Date

June 24, 1976

Acknowledgements

I would like to thank Professor Eberhard Wahl for his original suggestion to develop this model and for his continual encouragement and support. I am especially grateful to Professors Wayne Wendland and John Kutsbach for their patient instruction and valuable suggestions.

I am indebted to my parents for the efforts they made to interest me in knowledge throughout my childhood, and for their continued pride and interest in my achievements.

I owe special thanks to my wife Marian not only for encouragement, patience, and understanding but also for the gift of renewed awareness of the rewards of research. Her mother also deserves great appreciation for her professional and cheerful typing of this manuscript.

I am grateful to the United States Air Force for providing the opportunity and funding to make my studies and this research possible.

"In variation there is information."

- Reid Bryson

Table of Contents

	<u>Page</u>
List of Figures	iv
List of Tables	vii
Chapter 1 - Introduction	
1.1 Objectives.....	1
1.2 Previous analyses of the monsoon precipitation of India.....	3
Chapter 2 - Analysis and interpretation of data	
2.1 Establishment of data set.....	8
2.2 Normalization of data set.....	14
2.3 Fourier analysis of a single station's transformed data.....	19
2.4 Eigenvector analysis of spatial distribution of precipitation.....	23
Chapter 3 - Exploratory analysis of a method for the reconstruction of approximate precipitation records	
3.1 Evolution of model.....	40
3.2 Model verification within the data set for reduced numbers of stations.....	42
3.3 Model performance outside the data set for a reduced number of stations and reduced set of base eigenvectors.....	45
3.4 Substitution model for a severe drought year.....	54
Chapter 4 - Conclusions	
4.1 Results of this study.....	56
4.2 Recommendations for future research.....	58
Appendix I - Meteorological applications of variance analysis..	59
Appendix II - Values of components and coefficients for eigenvectors 1-22.....	64
References.....	74

List of Figures

	<u>Page</u>
2.1	Geographic location of stations listed in Table 2.1..... 11
2.2	Administrative subdivisions of the Indian subcontinent.... 12
2.3a	Histogram and equivalent normal distribution for measured and transformed April precipitation values, 1921-1960..... 15
2.3b	Same as 2.3a for July 1921-1960..... 16
2.3c	Same as 2.3a for October 1921-1960..... 17
2.4	Time series of the cube root transform of the reported July precipitation at Madras from 1921 to 1960..... 20
2.5	Line spectrum of transformed July precipitation record of Madras..... 21
2.6a	Components and coefficients of eigenvector 1 for transformed July precipitation values, 1921-1960..... 27
2.6b	Components and coefficients of eigenvector 2 for transformed July precipitation values, 1921-1960..... 28
2.6c	Components and coefficients of eigenvector 3 for transformed July precipitation values, 1921-1960..... 29
2.6d	Components and coefficients of eigenvector 4 for transformed July precipitation values, 1921-1960..... 30
2.6e	Components and coefficients of eigenvector 5 for transformed July precipitation values, 1921-1960..... 31
2.6f	Components and coefficients of eigenvector 6 for transformed July precipitation values, 1921-1960..... 31
2.6g	Components and coefficients of eigenvector 7 for transformed July precipitation values, 1921-1960..... 32
2.6h	Components and coefficients of eigenvector 8 for transformed July precipitation values, 1921-1960..... 32
2.6i	Components and coefficients of eigenvector 9 for transformed July precipitation values, 1921-1960..... 33
2.6j	Components and coefficients of eigenvector 10 for transformed July precipitation values, 1921-1960..... 33
2.6k	Components and coefficients of eigenvector 11 for transformed July precipitation values, 1921-1960..... 34

List of Figures (cont.)

	<u>Page</u>
2.6l Components and coefficients of eigenvector 12 for transformed July precipitation values, 1921-1960.....	34
2.6m Components and coefficients of eigenvector 13 for transformed July precipitation values, 1921-1960.....	35
2.7 Number of stations reporting excess or deficit of rainfall exceeding one transformed standard deviation from the transformed mean.....	38
3.1 Normalized departures of transformed values of measured precipitation for July 1921.....	43
3.2 July 1921 departures computed from the substitution model using 13 eigenvectors and 53 stations in the regression equation.....	43
3.3 July 1921 departures computed from the substitution model using 13 eigenvectors and 30 stations in the regression equation.....	43
3.4 July 1921 departures computed from the substitution model using 13 eigenvectors and 20 stations in the regression equation.....	43
3.5 Normalized departures for transformed values of measured precipitation for July 1920.....	46
3.6 July 1920 departures computed from the substitution model using 13 eigenvectors and 52 stations in the regression equation.....	46
3.7 July 1920 departures computed from the substitution model using 13 eigenvectors and 30 stations in the regression equation.....	46
3.8 July 1920 departures computed from the substitution model using 13 eigenvectors and 20 stations in the regression equation.....	46
3.9 July 1920 departures computed from the substitution model using 4 eigenvectors and 52 stations in the regression equation.....	49
3.10 July 1920 departures computed from the substitution model using 4 eigenvectors and 30 stations in the regression equation.....	49

List of Figures (cont.)

	<u>Page</u>
3.11 July 1920 departures computed from the substitution model using 4 eigenvectors and 20 stations in the regression equation.....	49
3.12 Normalized departures of transformed values of measured precipitation for July 1910.....	52
3.13 July 1910 departures computed from the substitution model using 13 eigenvectors and 52 stations in the regression equation.....	52
3.14 July 1910 departures computed from the substitution model using 13 eigenvectors and 30 stations in the regression equation.....	52
3.15 July 1910 departures computed from the substitution model using 13 eigenvectors and 20 stations in the regression equation.....	52
3.16 Normalized departures of transformed values of measured precipitation for July 1899.....	55
3.17 July 1899 departures computed from the substitution model using 13 eigenvectors and 47 stations in the regression equation.....	55
A.1 Observations of normalized temperature departures at stations A and B.....	61

List of Tables

	<u>Page</u>
2.1 Stations used in precipitation study.....	9
2.2 Chi square significance test.....	18
2.3 Results of Fourier analysis of Madras precipitation data....	22
2.4 Tabulation of eigenvalues, percent explained variance and cumulative percent explained variance for first 20 eigenvectors of transformed precipitation data.....	24
2.5 Total number of eigenvectors required to explain 80% of the variance contained in the transformed July precipita- tion records (1921-1960) for each of the stations.....	25
3.1 Comparison of model output to actual precipitation (July 1920) for stations not used in the 30-station regression equation.....	51

Chapter 1 - Introduction

Section 1.1 - Objectives

The vagaries of monsoon precipitation in the Indian subcontinent are a matter of common knowledge to both meteorologists and informed laymen. Climatically induced shortages of grain crops have brought the problem of analyzing and predicting monsoon rainfall to the attention of specialists in all fields of meteorology.

One of the obstacles for predictive models is the lack of a data set of sufficient density in either time or space. Time series must be sufficiently long to provide an initial data set for the development of the model and enough subsequent data to permit verification of the model. A complete time series provides the most reliable base for determining how specific sources of variation contribute to the total variance of a quantity. Also, "silent areas" (regions within the geographic domain of the predictive model with little or no available data) limit the model's effectiveness by limiting its spatial resolution.

The development of a model to provide approximate substitute meteorological records is the prime objective of this study. The July precipitation records of the Indian subcontinent were chosen as a test case for the substitution model since the results involve a region for which a predictive model is being prepared¹ and are based on records of such great variability as to provide convincing proof of the model's capabilities.

¹ The predictive model is being prepared by the Food-Climate group of the Institute for Environmental Studies at the University of Wisconsin-Madison. The model involves forecasts of agricultural yields and involves precipitation as only one of many inputs.

An efficient partitioning of the variance of the July monsoon precipitation records is the second objective of the study. The variance is examined through eigenvector analysis and the primary eigenvectors were used as the orthogonal basis for the substitution model.

Section 1.2 - Previous Analyses of the Monsoon Precipitation of India

One of the first systematic studies of the rainfall regimes of India was performed by H. F. Blanford (1889). This was the first descriptive analysis of the changing pattern of rainfall over the subcontinent. Blanford recognized that the variability of summer monsoon precipitation about the mean is greatest at those stations where the mean precipitation is smallest, and noted that the rainfall regimes exhibited a dependence on the surface wind flows associated with the "winter" and "summer" monsoons (monsoon is from the arabic mausim, meaning a time or a season, through the Middle Dutch monsoen). Blanford also instituted the practice of issuing forecasts of summer monsoon rainfall based on the timing and accumulation of spring snowfall on the western Himalayas.

Frequent failures of early forecasts led G. T. Walker (1923) to examine the concept of correlating worldwide meteorological events to summer monsoon rainfall over India. Choosing from among thousands of predictor-predictand equations those which had the highest correlation coefficients, Walker generated linear regression equations designed to forecast summer monsoon rainfall, a technique still used by the India Meteorology Department, though the predictors and the correlation coefficients have varied widely over the years. Among the predictors chosen by Walker and his successors have been the rainfall in Southern Rhodesia and Java, South American pressure, and wind persistence at Bangalore and Calcutta. Various investigators have since discovered that seasonal precipitation in India yields higher correlations with world meteorological events when used as a predictor rather than as a predictand; that is, the monsoon precipitation plays an active rather

than a passive role in world weather teleconnections (Normand, 1953). Nonetheless, the concept of isolating independent predictors of monsoon precipitation remains as an important contribution to monsoon forecasting techniques.

More recent authors have had the advantage of developments in statistical theory, digital computers and general advances in the field of theoretical climatology. Subrahmanyam (1969) analyzed the climatic regimes of India according to the 1955 water balance methodology of Thornthwaite and determined that considerable portions of the subcontinent utilized for food production fall under Thornthwaite's "dry subhumid" classification. According to Subrahmanyam, these areas experience total annual precipitation that in the mean just equals the water need, so even a slight departure from the mean results in a wide fluctuation of the water balance. Droughts in these areas are more common and more severe than in the other sections of the subcontinent and these regions could benefit the most from successful drought forecasting.

Jagannathan (1973) studied the annual rainfall records of 48 stations in India having record lengths of over 70 years to determine if any significant trends or oscillations occurred. By applying the Mann-Kendall rank statistic (whose significance test involves the null hypothesis that the time series is not different from a random series), he found that only nine of the records exhibited an increasing (linear or nonlinear) trend, while only two of the records exhibited a decreasing (linear or nonlinear) trend. These results are all significant on the 5% level. After subjecting all 48 series to a low pass filter (nine ordinates of the Gaussian probability curve) to suppress

the high frequency oscillations and a power spectrum analysis whose significance was tested by a null hypothesis using the red or white noise spectrum, he found that twelve stations exhibited significant (10% level or better) low frequency (greater than forty years) oscillations, eight exhibited a period of nearly eleven years, and twelve exhibited a quasi-biennial oscillation (periods from 2.0 to 3.2 years). Only three stations (Madras, Jagdalpur and Silchar) exhibited both cycles (11 year and QBO).

Moolley (1971) determined that there was little diurnal, semimonthly or monthly interdependence in summer rainfall over India, and that rainfall in the first half of the summer monsoon is independent of rainfall during the second half. Thus, excess or deficit of rainfall in the first half will not necessarily be balanced during the second half of the summer monsoon season. Such lack of dependence also dictates the failure of most predictive schemes based on the expectation that there is some level of persistence in the data base.

Several authors have investigated the degree of variability of the monsoon climate. Rao (1965) and Jagannathan (1973) used a "coefficient of variation" defined as the standard deviation expressed as a percentage of the mean, and found annual values ranging from 9% in Assam to 49% in Western Rajasthan. Similarly, Rao (1971) applied the Palmer Index function to India and determined how often the rainfall records indicated drought conditions. While those sections of India (west coastal India south of Bombay and northeastern India) that usually receive rainfall in excess of agricultural and water table replenishment seldom experience drought, other sections are plagued by drought up to 30% of the time during the summer monsoons.

Several authors have attempted to explain the great amount of variance in the monsoon by utilizing the techniques of harmonic and multivariate analysis. Lettau and White (1964) applied Fourier analysis techniques to the annual cycles of rainfall at 250 stations in India averaged over more than 50 years of observations. They found that the first three (of six) harmonics accounted for over 90% of the total variance, and that these three harmonics permitted a nearly complete assessment of the time and space development of the southwestern monsoon. It was possible to objectively study the onset date of the monsoon utilizing the phase angle of the harmonics to discover that the "peaking" of the monsoon occurs almost simultaneously from the west coast of India through the central plains to the Tibetan Plateau. The various rainfall regimes are well delineated by this technique which separates regions with summer maxima from areas with winter maxima and areas experiencing double maxima from those experiencing a single annual maximum.

Gangopadhyaya (1963) prepared a detailed study of five-day mean summer rainfall for a forty-year period for stations on the east and west coasts of India and discussed the characteristics of the patterns based on fitted orthogonal polynomials of the fifth degree. He determined the monsoon to be a complex composite of varied forcing functions, leading to the inherent "pulsatory" nature of the rainfall. He attributed this pulsatory nature to the passage of "easterly jet streams" across the central parts of the Bay of Bengal, peninsular India and the Arabian Sea.

While these harmonic analyses provide valuable objective statements of the nature of the averaged month to month variations of the monsoon,

they do not provide any insight into year to year variations. Use of other analysis techniques that interpret the interannual variance are required.

Jagannathan (1972) utilized eigenvector analysis to determine the orthogonal space fields of the monthly temperature patterns of peninsular India. These orthogonal fields represent the different predominant patterns of anomalies of temperature, with the first field (or pattern) explaining a greater amount of variance than the second field, and so on. Jagannathan determined that the first field accounted for as much as 71.1% of the variance present in the temperature data base for June, and only 26.7% of the August base. He found large intermonthly correlations for the first pattern, suggesting the possible existence of a common physical basis in the pattern throughout the year.

This review suggests the need for a similar study of the year to year variation in the precipitation of the monsoon. An objective analysis of the precipitation departure patterns and their interannual variability is a useful first step for developing a forecast technique. Such an examination may aid in the development of theories of monsoon mechanisms and the forecasting of the strength of these mechanisms.

Chapter 2 - Analysis and Interpretation of Data

Section 2.1 - Establishment of data set

Stations for the eigenvector analysis program were restricted to data archived on tape from the World Weather Record collection of meteorological data (Smithsonian Miscellaneous Collections, volumes 79, 90 and 105, and the World Weather Records of the U. S. Department of Commerce) as compiled by the National Center for Atmospheric Research. All available precipitation data for the Indian Subcontinent, including East and West Pakistan, India, the Laccadive and Andaman Islands, Sri Lanka, and Burma, were examined to determine which time period of at least thirty years had the most continuous precipitation records. The precipitation records of Burma and the Andaman Islands were excluded since no data was reported throughout the World War II era. A forty-year period was chosen so that a significant amount of the variability of the southwest monsoon would be included in the sample in the hope that the calculated eigenvectors would also be valid for years outside of the data set. The period 1921-1960 contained 53 stations with nearly continuous records. The stations are listed in Table 2.1 and their locations are shown in Figure 2.1. Figure 2.2 shows the general administrative subdivisions of the Indian subcontinent.

In this data set of 2120 monthly precipitation values, 52 points (about 2.5%) were missing from the records. Since the eigenvector analysis program would treat the missing values as zeroes, appropriate substitutions were required. Ordinarily, a ten- or twenty-year mean value is substituted, but it was thought this would not be representative in a set of records as variable as these. Accordingly, the data of 1921-1960 for the missing station and three neighboring

Table 2.1. Stations used in precipitation study. Mean and standard deviation values listed are for the appropriate cube root values discussed in Section 2.2.

ID number	Station	W.M.O. number	Elevations (meters)	July average cube root precipitation ($\text{cm}^{0.33}$)	Standard deviation of July cube root precipitation ($\text{cm}^{0.33}$)
1	Peshawar	41530	359	1.323	.591
2	Lahore	41640	214	2.250	.486
3	Siala	42083	2202	3.437	.381
4	Quetta	41661	1673	1.120	.566
5	Kelat	41696	2017	1.220	.597
6	Ludhiana	42099	247	2.677	.417
7	Mukteswar	42147	2311	3.142	.416
8	Bikaner	42165	224	1.955	.613
9	Agra	42261	169	2.678	.465
10	Darjelling	42295	2128	4.146	.354
11	Dibrugarh	42312	106	3.711	.391
12	Karachi	41782	4	1.789	.795
13	Jodhpur	42339	224	2.188	.542
14	Jaipur	42348	390	2.597	.438
15	Darbhanga	42391	49	3.120	.514
16	Dhubri	42404	35	3.457	.616
17	Gauhati	42411	55	3.063	.355
18	Kota	42451	257	3.008	.552
19	Allahabad	42475	98	3.076	.444
20	Patna	42491	53	2.977	.358
21	Cherrapunji	42515	1313	6.140	.787
22	Shillong	42516	1500	3.257	.521
23	Daltonganj	42587	149	3.241	.392
24	Dumka	42599	149	3.323	.351
25	Silchar	42619	29	3.746	.371
26	Sagar	42671	551	3.519	.509
27	Dwarka	42731	11	2.505	.800
28	Indore	42754	567	3.068	.464

Table 2.1 (continued)

ID number	Station	W.M.O. number	Elevations (meters)	July average cube root precipitation (cm ^{0.33})	Standard deviation of July cube root precipitation (cm ^{0.33})
29	Calcutta	42807	6	3.139	.402
30	Nagpur	42867	310	3.364	.362
31	Veraval	42909	8	2.797	.779
32	Akola	42933	282	2.854	.416
33	Cuttack	42970	27	3.242	.409
34	Jagdalpur	43041	553	3.371	.437
35	Bombay	43057	11	4.016	.560
36	Poona	43063	559	2.517	.440
37	Begampet	43128	545	2.520	.324
38	Vishakhapat	43149	3	2.191	.397
39	Masulipatam	43185	3	2.601	.417
40	Belsaum	43197	753	3.625	.416
41	Madras	43279	16	1.947	.384
42	Mangalore	43283	22	4.692	.435
43	Bangalore	43295	921	2.187	.363
44	Amini	43311	4	2.535	.971
45	Kodaikanal	43339	2343	2.190	.403
46	Fort Cochin	43351	3	3.819	.488
47	Pamban	43363	11	.990	.376
48	Minicoy	43369	2	2.763	.442
49	Trivandrum	43371	64	2.731	.481
50	Trincomalee	43418	7	1.570	.603
51	Colombo	43466	6	2.268	.637
52	Nuwaraeliya	43473	1880	2.816	.473
53	Hambantota	43497	20	1.492	.592

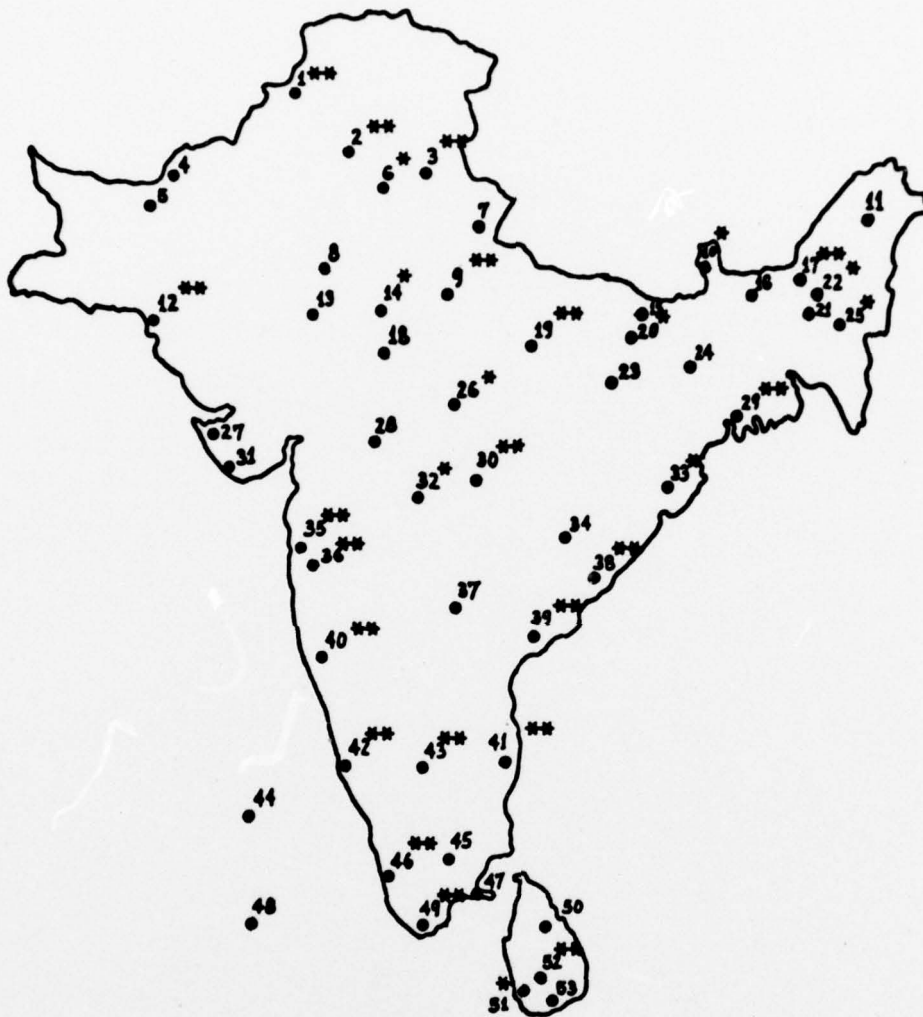


Figure 2.1. Geographic location of stations listed in Table 2.1. Stations marked with either a single or a double star are among the 30 stations with the longest precipitation records. Stations marked with a double star are among the 20 stations with the longest precipitation records.

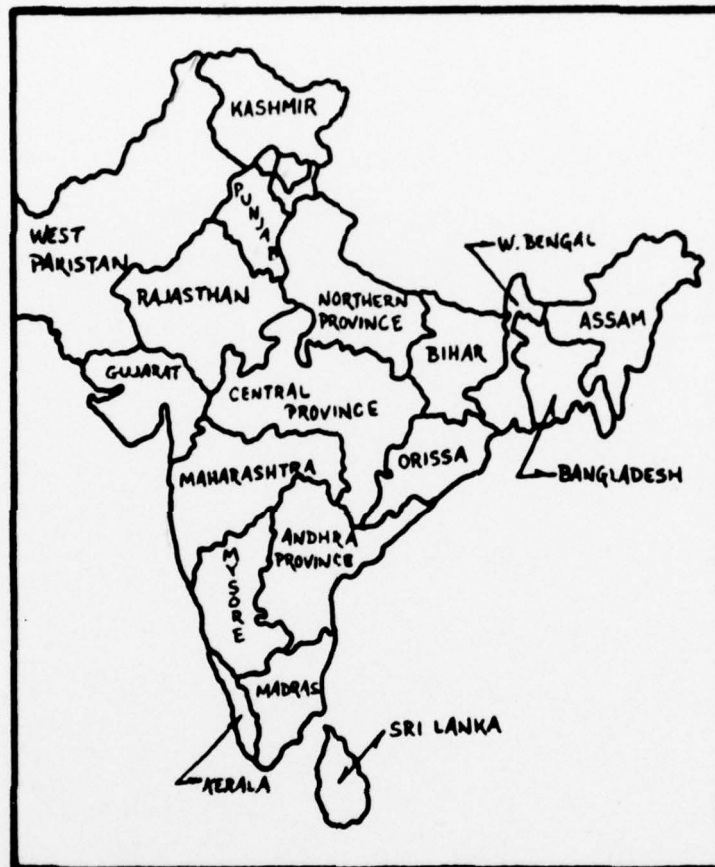


Figure 2.2. Administrative subdivisions of the Indian subcontinent.

stations was rank ordered and identified by year. For the year of missing data, the rank in the neighboring three stations was determined. The average rank identified a value in the original, data deficient station record. This value and the next largest and smallest values were averaged and the result used to substitute for the missing value.

During the data analysis, it was noted that the data listed for Amini Divi in the Laccadive Islands for the period 1941-1950 was incorrectly labelled. While the data listed on page 788 of the 1941-1950 Collection of World Weather Records are actually given in inches of precipitation, they were transferred with the same values but the units were listed as millimeters. The corrected numbers were substituted into the data set for this analysis.

Section 2.2 - Normalization of Data Set

Since most statistical tests of significance are based on the premise that the tested sample comes from a normal population, a mathematical conversion was used to normalize the precipitation data set. Moolley (1973), in an analysis of monthly rainfall records of India, showed that in 70% of the cases examined the cube root transformation led to normalization, while the performance of the logarithmic transformation was poor. Also, the logarithm of zero is $-\infty$, so the logarithmic transformation is rather awkward for reports of zero precipitation. Either the zero precipitation observation must be replaced by a small but finite amount, or the problem must be circumvented by substituting a value (usually 0.0) for the transformed value. For some months zero precipitation is the modal value, so use of the logarithmic transformation would have been unsatisfactory.

A single station test was performed on the precipitation record of Mangalore comparing the appropriate normal curve, based on the mean and standard deviation, to the histograms of the measured precipitation and the logarithmic and cube root transformation. The April, July, and October records were analyzed since these months represent transition months and a typical monsoon month. Figure 2.3a, b, and c shows the histograms for April, July, and October over which have been superimposed the appropriate normal curves. Although no transformation is perfect, it appears that the cube root transformation produces a more normal data set than does the logarithmic transformation.

A chi square test was applied to all nine analyses with the null hypothesis that the observed distribution was not significantly

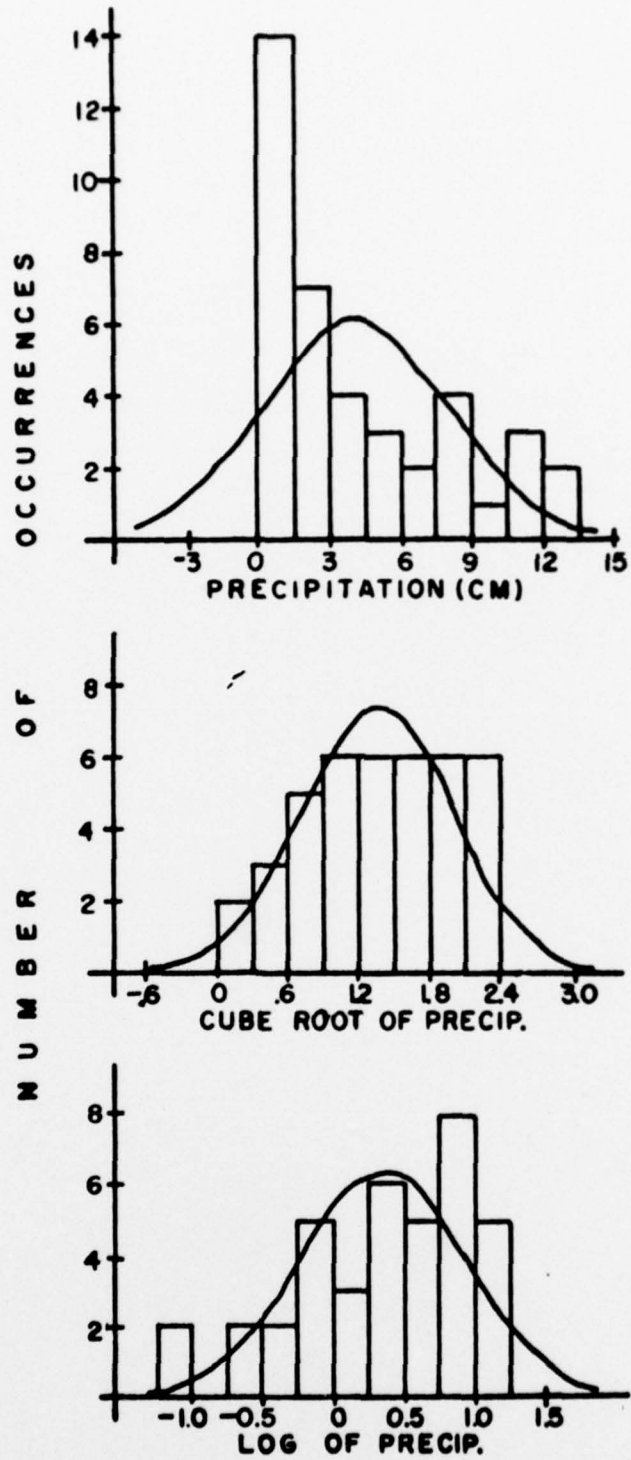


Figure 2.3a. Histogram and equivalent normal distribution for measured and transformed April precipitation values, 1921-1960.

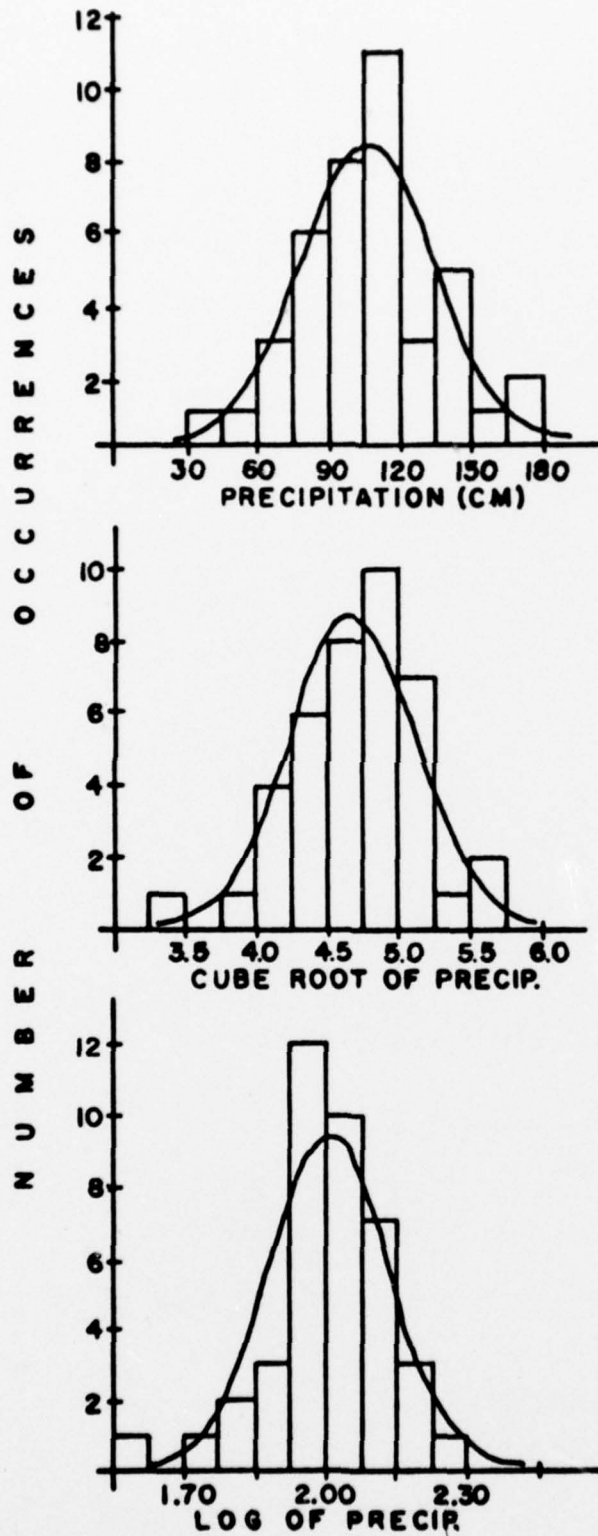


Figure 2.3b. Same as 2.3a for July 1921-1960.

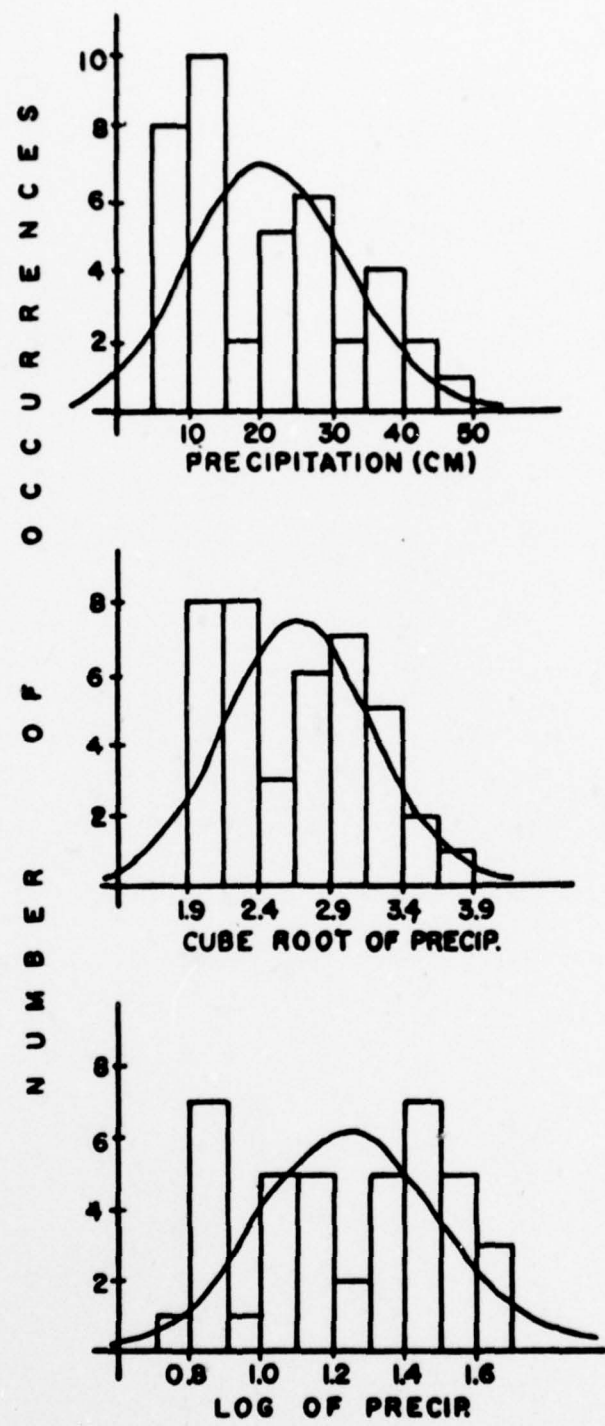


Figure 2.3c. Same as 2.3a for October 1921-1960.

different from the normal distribution. The results are tabulated in Table 2.2. The chi square value for each of the three months was smallest for the cube root transformation, and the probability of obtaining the sample chi square from a normal population for the April, July and October samples was greater than 95%, 95%, and 70%. The data set was transformed to cube root values for all further calculations. Through the rest of this study, the word "transformed" indicates the appropriate cube root value.

Table 2.2. Chi square significance test. Significance level exceeded by each distribution based on null hypothesis that the observed distribution was not significantly different from the normal distribution.

<u>Month</u>	<u>Significance level exceeded</u>		
	<u>Measured precipitation</u>	<u>Cube root transformation</u>	<u>Logarithmic transformation</u>
April	.50	.95	.70
July	.90	.95	.80
October	.20	.70	.30

Section 2.3 - Fourier Analysis of a Single Station's Transformed Data

The 1921-1960 July transformed precipitation record for Madras was Fourier analyzed to identify any dominant frequencies in the time series. Figure 2.4 shows the time series of the transformed data which shows large annual fluctuations and few apparent trends. Figure 2.5 shows the line spectrum of the twenty components of the harmonic analysis. No attempt was made to determine the significance of the components. Only the harmonic for .3 cycles per year (twelve cycles per forty years) appears to have a considerably greater amplitude than the others. There is little evidence for an eleven-year cycle, and the components associated with a possible QBO are no greater than at least four other harmonics of the set.

As shown in Table 2.3, Fourier analysis required fourteen harmonics (70% of the possible harmonics) to explain 80% of the variance. There seems to be little chance of identifying the rainfall pattern with a dominant harmonic.

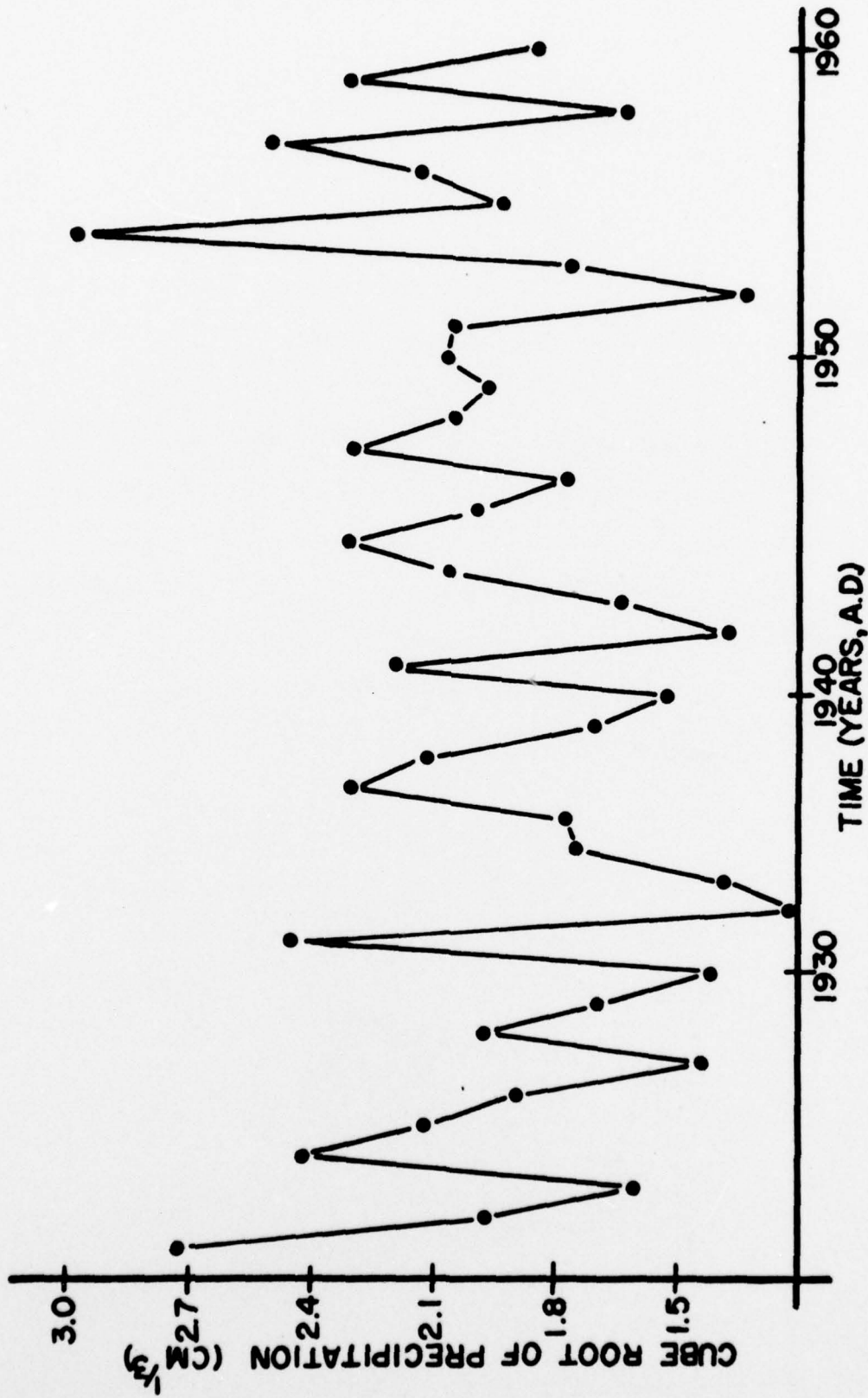


Figure 2.4. Time series of the cube root transform of the reported July precipitation at Madras from 1921 to 1960.

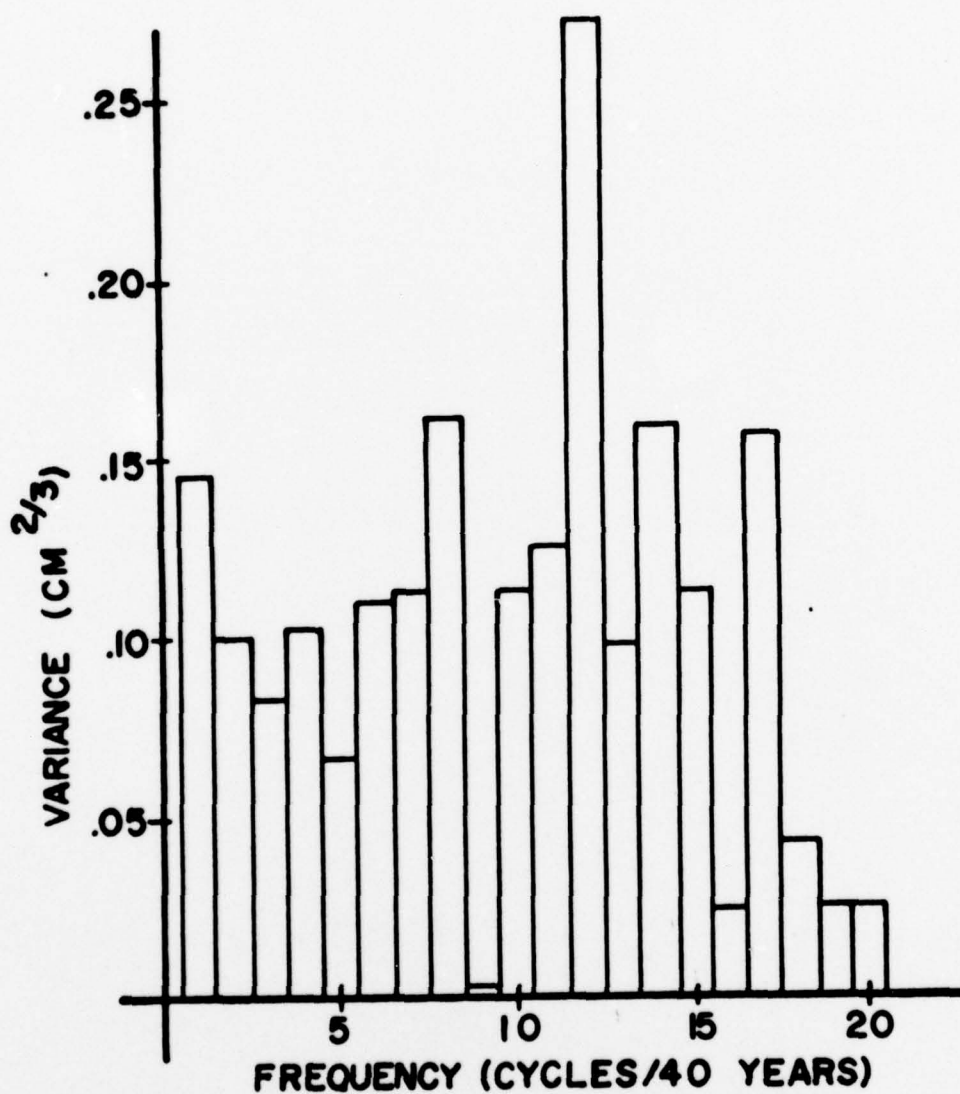


Figure 2.5. Line spectrum of transformed July precipitation record of Madras. The variance of each harmonic equals $\frac{1}{4}$ the square of the amplitude for that harmonic, except for the twentieth harmonic where the variance equals the amplitude squared.

Table 2.3. Results of Fourier analysis of Madras precipitation data.

<u>Harmonic</u>	<u>Amplitude</u>	<u>Variance explained by "I"th Harmonic (%)</u>	<u>Cumulative explained variance (%)</u>
1	0.196	12.6	12.6
2	0.100	3.3	15.9
3	0.081	2.2	18.1
4	0.107	3.8	21.9
5	0.067	1.5	23.4
6	0.111	4.0	27.4
7	0.112	4.1	31.5
8	0.163	8.8	40.3
9	0.007	0.0	40.3
10	0.111	4.1	44.4
11	0.128	5.4	49.8
12	0.273	24.6	74.4
13	0.100	3.3	77.7
14	0.160	8.4	86.1
15	0.114	4.3	90.4
16	0.027	0.2	90.6
17	0.158	8.3	98.9
18	0.042	0.5	99.4
19	0.028	0.3	99.7
20	0.028	0.3	100.0

Section 2.4 - Eigenvector Analysis of Spatial Distribution of Precipitation

For those readers who have not previously encountered eigenvector analysis, a qualitative discussion of the physical meaning of eigenvectors is included in Appendix I.

The transformed records for each station were normalized by dividing the transformed value minus the transformed mean by the forty-year transformed standard deviation for that station, so the variance for each of the 53 stations equalled 1.0 and the total variance equalled 53. The data forms a 53 station by 40 year matrix whose correlation matrix has 53 rows and columns. Eigenvalues, eigenvectors, and coefficients of the eigenvectors were computed following the method of Kutzbach (1967). The components of the first twenty-two eigenvectors and their coefficients for the years 1921-1960 are tabulated in Appendix II. For computational efficiency, the 53 by 53 matrix was transformed to a 40 by 40 matrix for the eigenvector calculations following the method of Hirose and Kutzbach (1969). The 40-component eigenvectors were then transformed into 53-component eigenvectors and it is these eigenvectors of the spatial patterns of transformed precipitation that will now be described.

Table 2.4 lists the eigenvalues and percent explained variance for the first twenty eigenvectors. The first eigenvector explains 17% of the variance; the first six eigenvectors explain over 50% of the variance; the first twenty eigenvectors (50% of the total orthogonal set) explain over 90% of the variance. Table 2.5 shows the total number of eigenvectors required to explain 80% of the variance contained in the transformed July precipitation records for each of the stations (the stations are identified by number).

Table 2.4. Tabulation of eigenvalues, percent explained variance and cumulative percent explained variance for first 20 eigenvectors of transformed precipitation data.

Eigenvector	Eigenvalue	Explained variance (%)	Cumulative explained variance (%)
1	9.00	17.0	17.0
2	6.33	11.9	28.9
3	4.16	7.9	36.8
4	3.65	6.9	43.7
5	2.76	5.2	48.9
6	2.55	4.8	53.7
7	2.26	4.3	58.0
8	2.15	4.1	62.1
9	1.97	3.7	65.8
10	1.63	3.4	69.2
11	1.51	3.1	72.3
12	1.39	2.9	75.2
13	1.20	2.6	77.8
14	1.14	2.3	80.1
15	1.06	2.2	82.3
16	0.90	2.0	84.3
17	0.83	1.7	86.0
18	0.73	1.5	87.5
19	0.69	1.4	88.9
20	0.64	1.3	90.2

Table 2.5. Total number of eigenvectors required to explain 80% of the variance contained in the transformed July precipitation records (1921-1960) for each of the stations (for station identity, see Table 2.1).

Number of eigenvectors	Stations
7	21, 49
8	-
9	46
10	5, 14, 32, 43
11	4, 16, 35, 36, 51
12	3, 30, 45
13	1, 25, 26, 31, 47, 53
14	9, 12, 13, 18, 27, 38, 40, 42, 48, 52
15	2, 6, 11, 23, 24, 39
16	8, 15, 17, 22, 28, 29, 33
17	41, 44, 50
18	19, 34
19	10
20	7, 37
21	20

If the spatial distribution of variance were entirely random, each of the eigenvectors could be expected to explain an equal percentage of the variance. For forty eigenvectors, this would be 2.5% of the variance. Since the final 27 eigenvectors each explain less than 2.5% of the variance, they are considered to carry too little information to warrant further investigation. The remaining thirteen eigenvectors (which together explain 78% of the variance) and the time series of their coefficients are shown in Figures 2.6a through 2.6m.

Eigenvectors one through four explain 44% of the total variance. All four patterns are characterized by large, coherent departure areas separated by distinct, strong gradients. This implies that these regions have distinctly different rainfall regimes, and when one area is experiencing drought, a neighboring region may be experiencing floods. Proof that this is the case in India is borne out by the historical records of floods and famine as well as rainfall departure maps for individual years.

Positive and negative precipitation departures (eigenvector component values) alternate in the north-south direction in eigenvectors 1 and 4 (which explain 24% of the variance) and in the east-west direction in eigenvectors 2 and 3 (which explain 20% of the variance). This gives some credence to the belief that the northward or southward displacement of the monsoon trough is a slightly more important factor in the monsoon precipitation than is its eastward or westward position. However, since many meteorological factors are involved and the difference is slight, more evidence is certainly required.

Eigenvector 1 shows three distinct regions of precipitation departures. This "most preferred pattern of variability" indicates

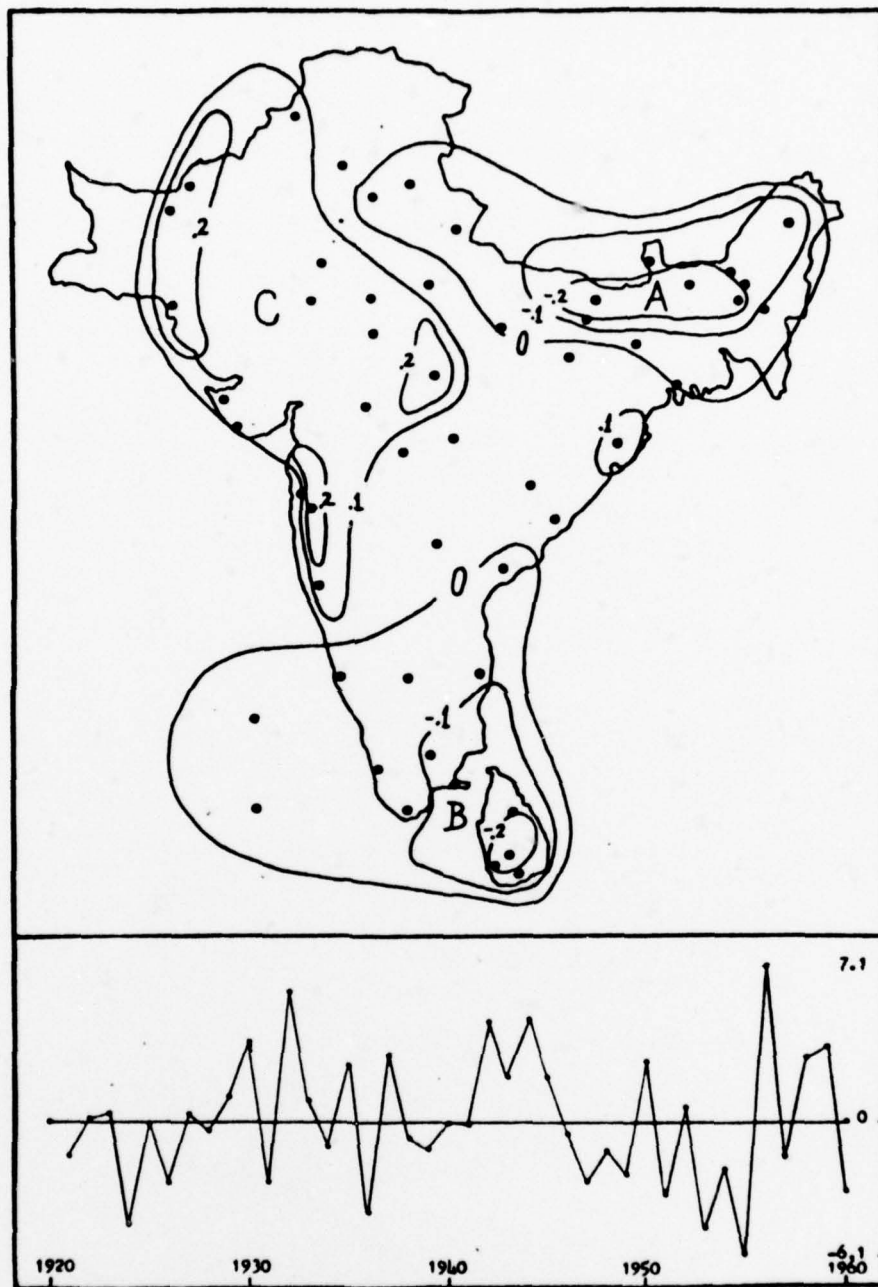


Figure 2.6a. Components and coefficients of eigenvector 1 for transformed July precipitation values, 1921-1960.

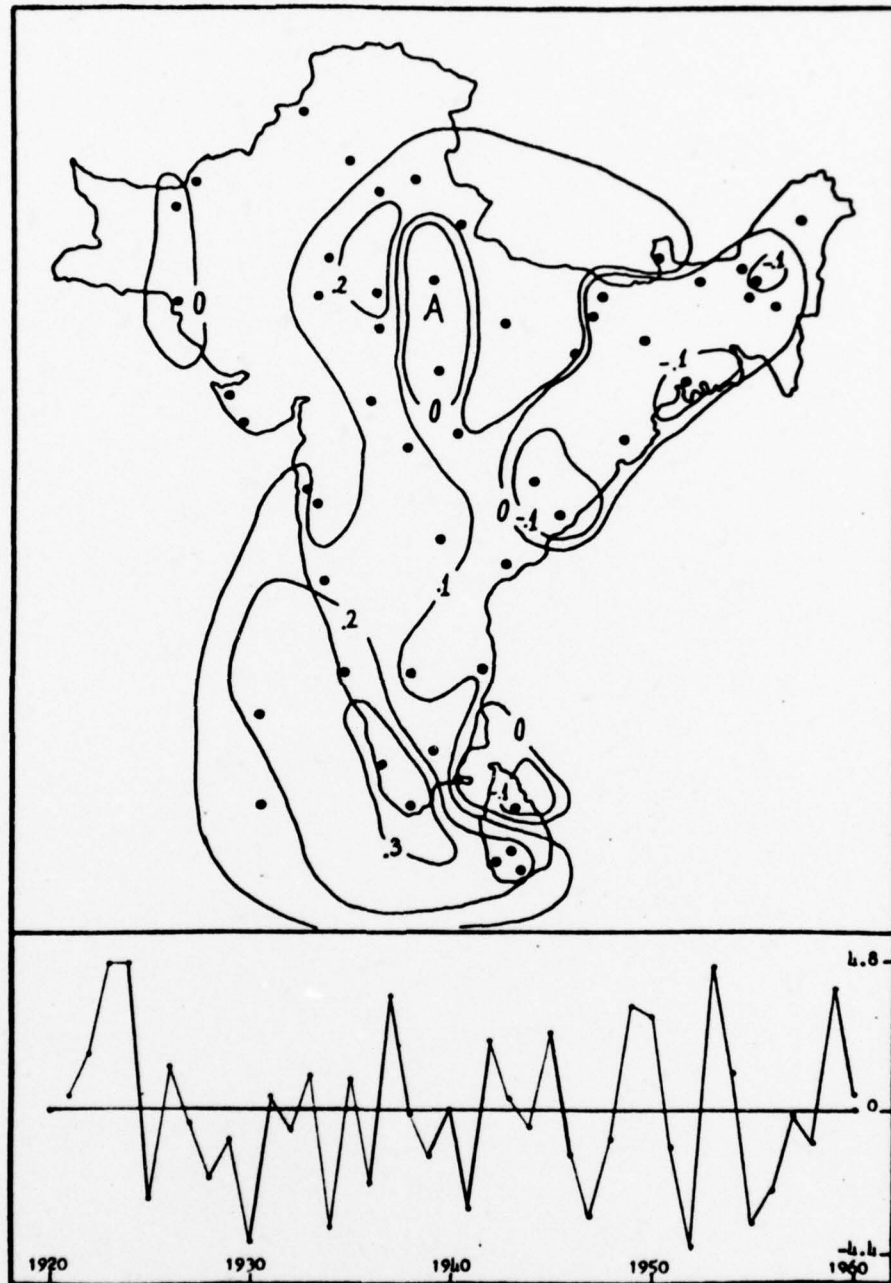


Figure 2.6b. Components and coefficients of eigenvector 2 for transformed July precipitation values, 1921-1960.

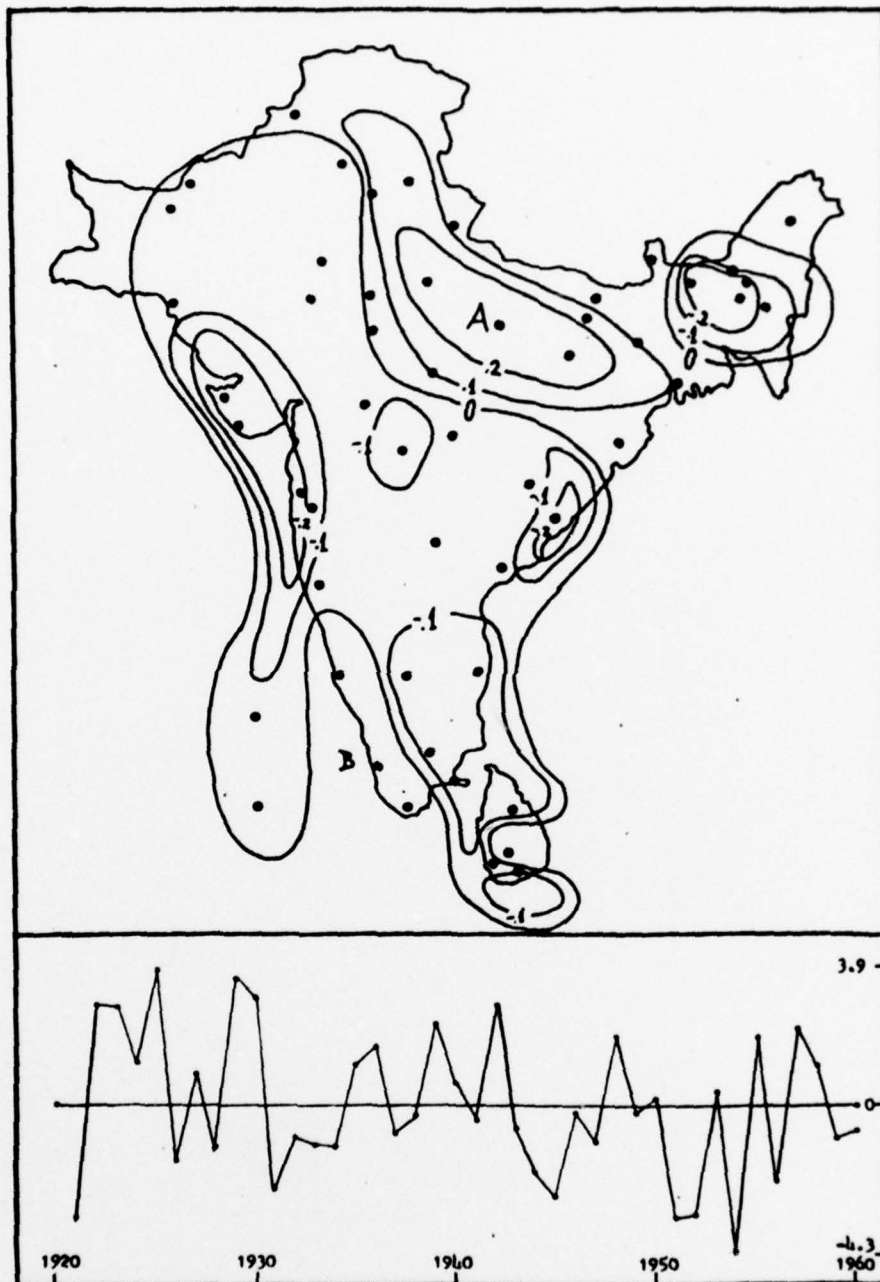


Figure 2.6c. Components and coefficients of eigenvector 3 for transformed July precipitation values, 1921-1960.



Figure 2.6d. Components and coefficients of eigenvector 4 for transformed July precipitation values, 1921-1960.

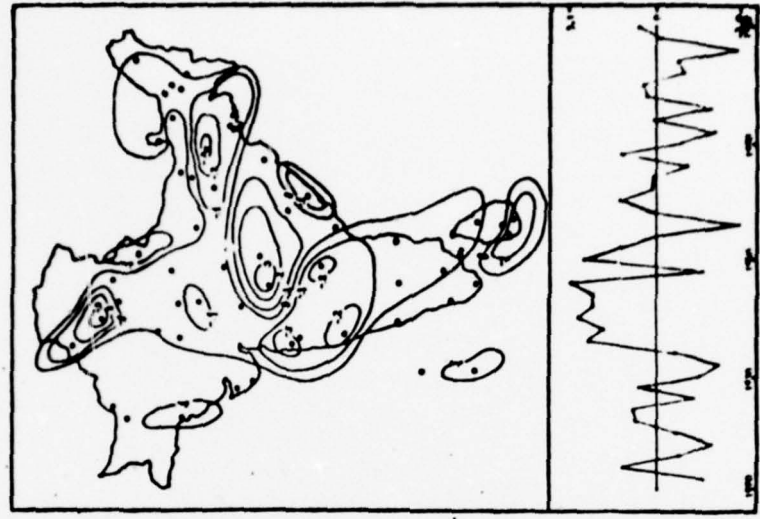


Figure 2.6f. Components and coefficients of eigenvector 6 for transformed July precipitation values, 1921-1960.

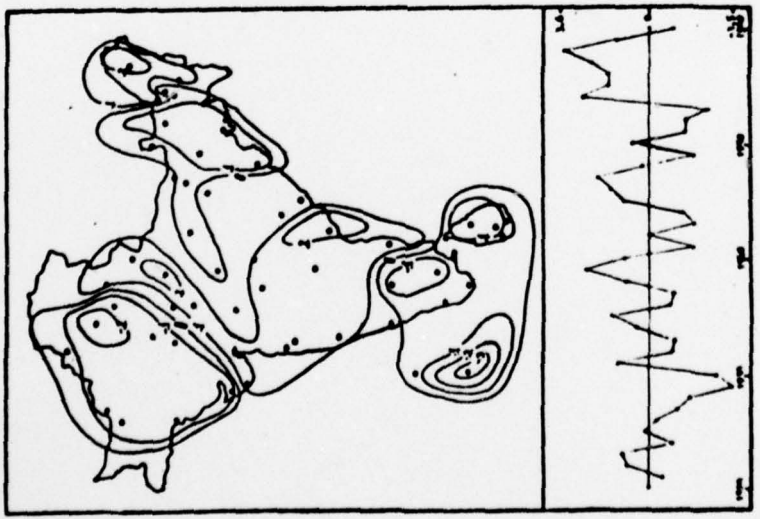


Figure 2.6e. Components and coefficients of eigenvector 5 for transformed July precipitation values, 1921-1960.

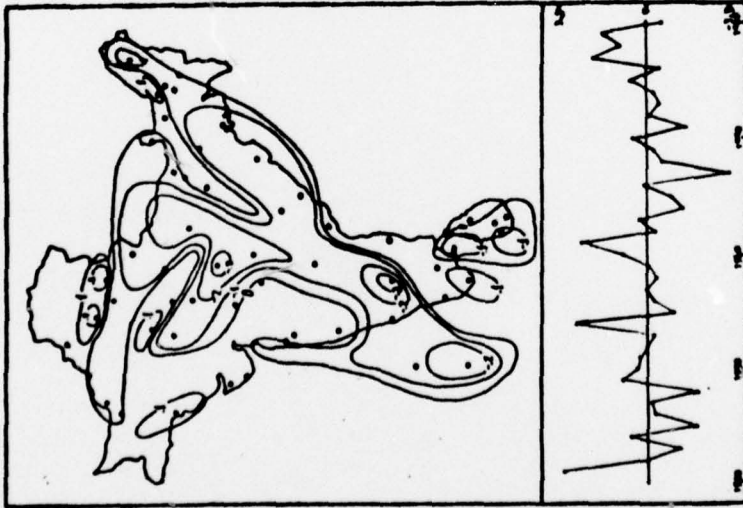


Figure 2.6g. Components and coefficients of eigenvector 7 for transformed July precipitation values, 1921-1960.

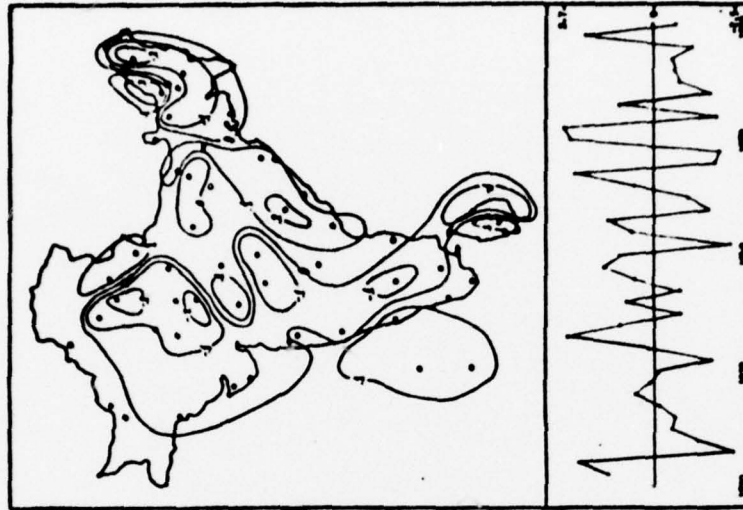


Figure 2.6h. Components and coefficients of eigenvector 8 for transformed July precipitation values, 1921-1960.



Figure 2.6i. Components and coefficients of eigenvector 9 for transformed July precipitation values, 1921-1960.

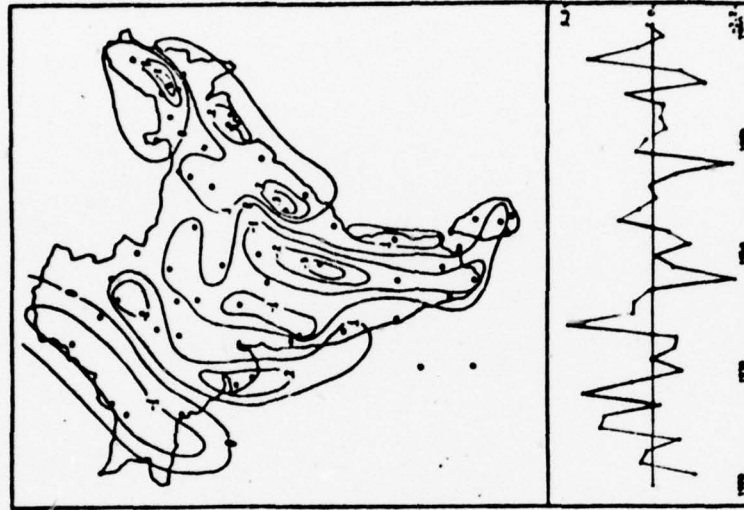


Figure 2.6j. Components and coefficients of eigenvector 10 for transformed July precipitation values, 1921-1960.

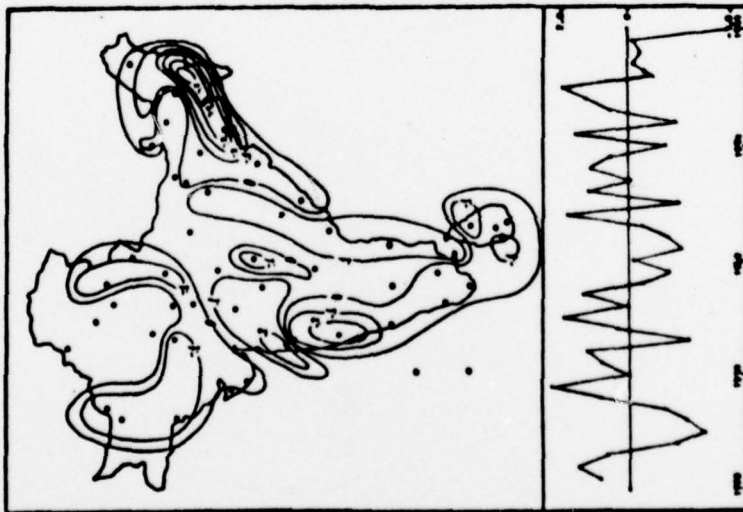


Figure 2.6k. Components and coefficients of eigenvector 11 for transformed July precipitation values, 1921-1960.

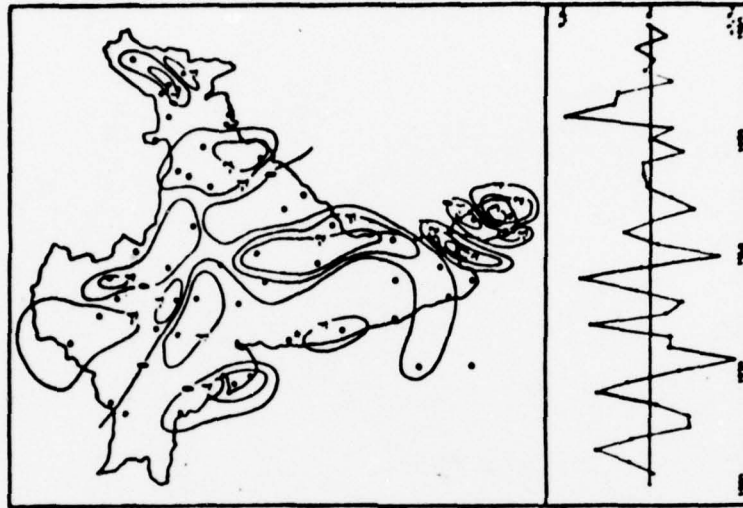


Figure 2.6l. Components and coefficients of eigenvector 12 for transformed July precipitation values, 1921-1960.

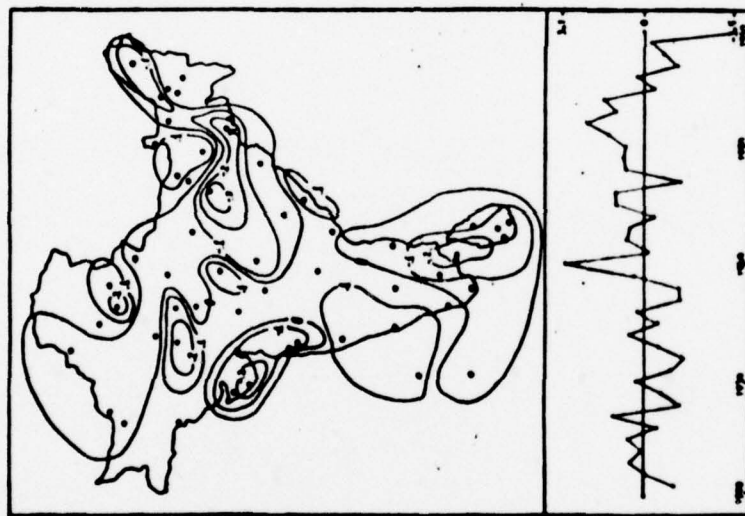


Figure 2.6a. Components and coefficients of eigenvector 13 for transformed July precipitation values, 1921-1960.

negative components from Assam through Bihar into Punjab province (region A in Figure 2.6a), and in the southeastern peninsular region (Andhra, Mysore, Kerala and Madras provinces) including the Laccadive Islands and Sri Lanka (B in Figure 2.6a). Positive components are associated with the stations of northwestern India and West Pakistan (C in Figure 2.6a). The eigenvector coefficients show that this pattern was evident in 18 of the 40 years in the record, and was strongly present in 1932, 1942, 1944, and 1956. The negative of this pattern was evident in the remaining 22 years of the 40-year record and was strongly present in 1924, 1936, 1947, 1955, and 1960. It is particularly interesting to note that the two extreme coefficients occurred in consecutive years (1955 and 1956), which is continuing evidence of the highly variable nature of the precipitation patterns.

Eigenvector 4 also shows precipitation departures that alternate in the north-south direction, but in contrast to the positioning of the departures in eigenvector 1, the dominant positive zone runs from Bombay through Calcutta (A in Figure 2.6d) with a small positive region in extreme southern India and southern Sri Lanka (B in Figure 2.6d). This pattern was most evident in 1923, 1932, 1941, and 1951 while its negative image was most evident in 1936 and 1956. Eigenvector 3 also has a dominant positive zone (A in Figure 2.6c) from Kashmir through Orissa, with a small positive region (B in Figure 2.6c) along the west coast, mostly in Kerala Province. This pattern was most evident in 1925, 1929, and 1942, while its negative image was most evident in 1921, 1931, and 1954.

Eigenvector 2 shows primarily negative precipitation departures along the east coast, and positive departures along the west coast

and through the north central region of the country with the exception of a narrow negative zone in the Northern Province (A in Figure 2.6b). Eigenvector 2 was most evident in 1923, 1924, 1937, 1953, and 1959 while its negative image was most evident in 1930, 1934, 1941, 1947, 1952, and 1956.

The patterns for eigenvector five through thirteen are more complex than the four previous patterns and offer no discernible physical interpretation, due to a blending of the local effects and the remaining macroscale patterns. The local effects may represent the adjustment of the macroscale circulation pattern to micro or mesoscale geomorphic features, or they may only be nonrepresentative values of a certain station's record. In either case, the presence of the local effects masks the macroscale pattern represented by the eigenvector. Thus, although eigenvectors one through thirteen present some of the macroscale pattern of transformed precipitation, eigenvectors one through four appear most likely to do so without much interference of local effects.

As an indicator of years with extreme precipitation, the number of stations from the 53-station network reporting July precipitation greater than one transformed deviation above or below the transformed mean were counted for each year during the 1921-1960 period. Since these conditions are exceeded only 32% of the time, the magnitude of the number of stations reporting this departure from normal precipitation is interpreted as a measure of severity for agricultural purposes, since it is likely that grain production would be greatly affected by this much excess or deficit rainfall. The results are graphed in Figure 2.7. The years 1923, 1924, 1937, 1953, and 1959 showed unusually

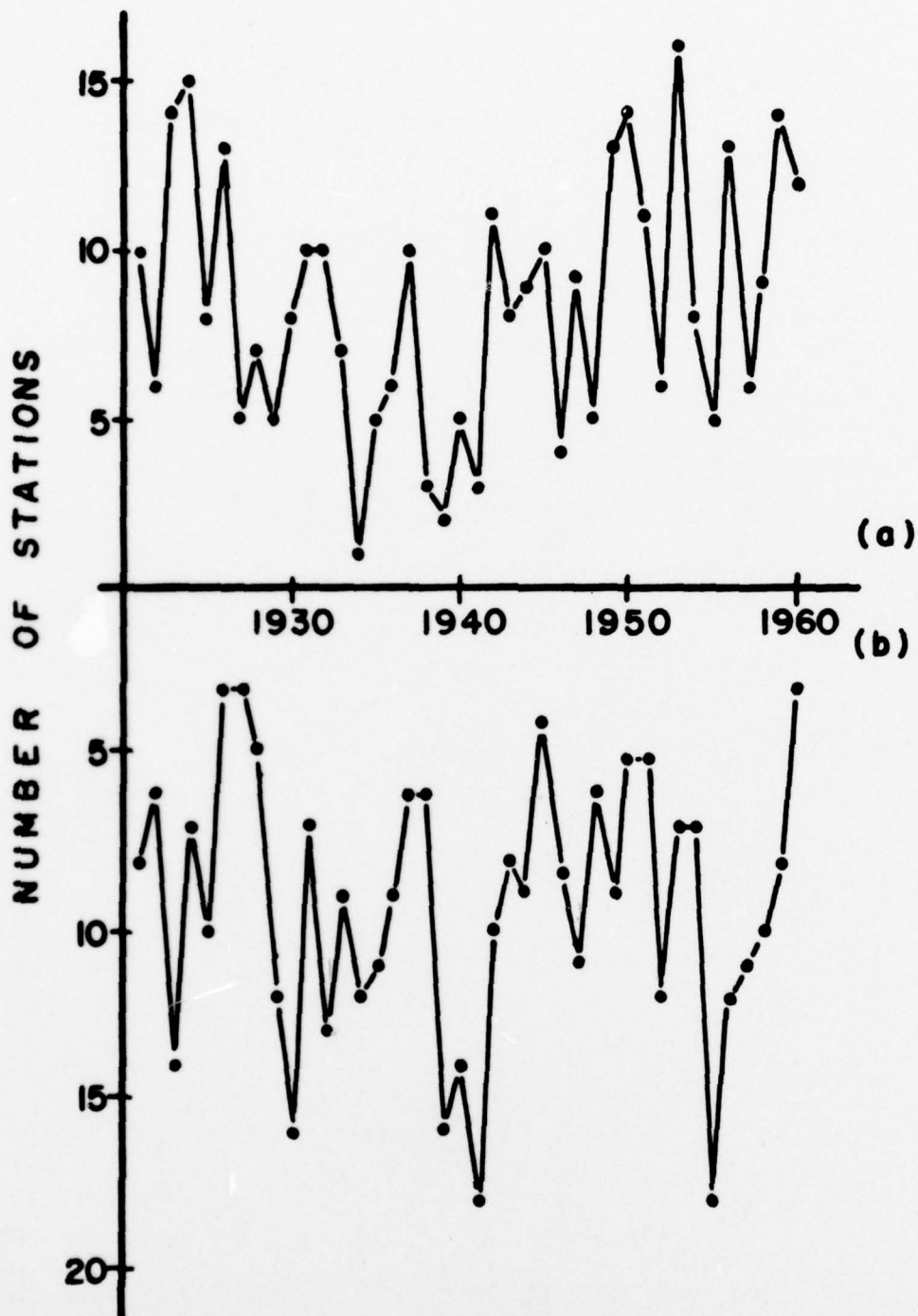


Figure 2.7. Number of stations reporting excess or deficit of rainfall exceeding one transformed standard deviation from the transformed mean. Stations reporting excess are plotted in (a), stations reporting deficit are plotted in (b).

high numbers of stations reporting excessive precipitation, while 1930, 1941, and 1955 showed an unusually high number of stations reporting deficient rainfall. These were years for which the eigenvector coefficients for eigenvector 2 had high and low values respectively. The correlation between the eigenvector coefficients and the number of stations with extremely high precipitation minus the number of stations with extremely low precipitation was .66 which indicates that at least 44% of the variance of the number of stations with extreme precipitation can be accounted for by the coefficient of the second eigenvector. A similar correlation with the eigenvector coefficients for eigenvectors 1, 3 and 4 produced correlation coefficients of -0.06, -0.34, and -0.31 (0.3%, 11% and 9% of the variance) respectively. Thus it appears that the second eigenvector best portrays the variance pattern that occurs when the monsoon rains are unusually strong for many regions, while its negative image occurs when the monsoon rains are unusually weak.

Chapter 3 - Exploratory Analysis of a Method for the Reconstruction
of Approximate Precipitation Records

Section 3.1 - Evolution of Model

The computed eigenvectors analyzed in section 2.4 provide a partial representation of all the transformed precipitation departures for the years 1921-1960. The total departure for any station will be the linear combination of the components of all forty eigenvectors and the coefficients of each eigenvector, but the departure can be approximated by truncating this series. Since the eigenvectors are orthogonal, the coefficients could be determined from a multiple linear regression between the known departures and eigenvectors. Within the data set, these coefficients will be nearly identical to those computed as part of the eigenvector analysis (they would be exactly identical if all 40 eigenvectors were used instead of the truncated series).

Of the 53 stations whose records for 1921-1960 were either complete or nearly so, only 40 have nearly continuous records to 1891, 30 to 1870, and 10 to 1856. The missing data for these years can be approximated from the reported data so long as we assume that the components for each eigenvector for each station remains the same as that found for the period 1921-1960; that is, we assume that the eigenvector patterns do not change with time. The known departures can be expressed as a linear combination of N known eigenvectors,

let: D_i^{aa} = the normalized departure for the i^{th} station for the year aa

C_i^{aa} = the coefficient for the I^{th} eigenvector for the year aa

$E_{I,i}$ = the component of the I^{th} eigenvector at the i^{th} station

then $D_i^{aa} = C_1^{aa} \cdot E_{1,i} + C_2^{aa} \cdot E_{2,i} + \dots + C_N^{aa} \cdot E_{N,i}$.

The coefficients C can be obtained from the multiple linear regression between the known departures D and known eigenvector components E. These coefficients can be used to calculate departures for any station, as long as the eigenvector components are known for that station from the original forty-year data analysis, by taking the coefficient times the component for that station, summed over all of the eigenvectors.

This model must be evaluated for a number of years to check several different assumptions. The assumption that the eigenvector patterns do not change with time can be assessed indirectly by determining if the accuracy of the model seriously deteriorates with time. The sensitivity of the model to the number of eigenvectors used in the regression can be evaluated by comparing the results obtained for a single year when more or fewer eigenvectors are used in the regression. The sensitivity of the model to the number of stations used in the regression can be found by comparing the model output for progressively fewer stations within a single year.

It must be remembered that with the limited number of eigenvectors, the model will be able to explain only a portion of the variance of the data set. Even within the original 40-year analysis period, the set of 13 eigenvectors explained only 78% of the variance. Since the eigenvectors are expected to identify generalized rather than localized patterns of variability, the model will seldom reconstruct the extreme maxima and minima of a given year's departures. At best, it will provide a record of general trends for years in which no better estimate is available.

Section 3.2 - Model Verification within the Data Set for Reduced Numbers of Stations

The sensitivity of the model to the number of stations included in the regression was first tested within the original data set by analyzing the data for 1921. The observed normalized departure pattern using the transformed precipitation data for 1921 is shown in Figure 3.1. July precipitation was deficient over much of the subcontinent except for most of Andhra and Madras Provinces, northern Assam Province, the Bombay-Poona area, and Sri Lanka. The largest departures were -2.37σ (12.3 cm) at Sagar (A in Figure 3.1) and 2.02σ (20.2 cm) at Madras (B in Figure 3.1). A narrow zone of excess rainfall (C in Figure 3.1) extended from Lahore through Bikaner.

The first thirteen eigenvectors discussed in Section 2.4 were used as the truncated orthonormal base for the regression. Stations were selected for inclusion in the regression on the basis of length of record. Calculations were performed using 52, 30 and 20 stations in the regression. Locations of the stations used in each regression are shown in Figure 2.1.

The substitute coefficients obtained by regressing the departures at 52 stations against the eigenvectors were nearly identical to the eigenvector coefficients from the original computation--none deviated from the original by more than .1%. Eigenvectors 3 and 7 were the dominant patterns. The departures calculated from combinations of the coefficients and eigenvectors are shown in Figure 3.2. Since the substitute and original coefficients are identical, the figure shows both the pattern of departures generated by the original analysis and the pattern generated by the substitution model. The reconstructed pattern is quite similar to the observed departure pattern. The



Fig. 3.1. Normalized departures of transformed values of measured precipitation for July 1921.



Fig. 3.2. July 1921 departures computed from the substitution model using 13 eigenvectors and 53 stations in the regression equation.



Fig. 3.3. July 1921 departures computed from the substitution model using 13 eigenvectors and 30 stations in the regression equation.

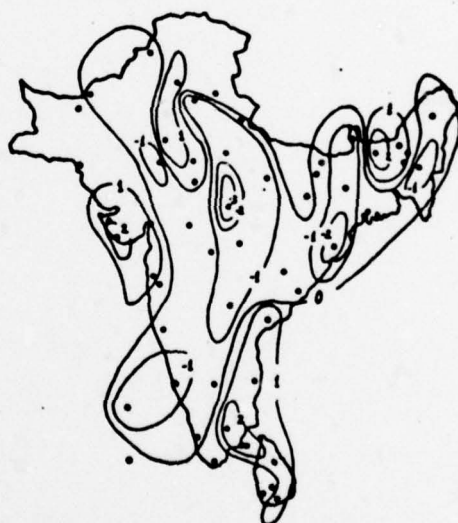


Fig. 3.4. July 1921 departures computed from the substitution model using 13 eigenvectors and 20 stations in the regression equation.

correlation between the observed and reconstructed records is + 0.89. The deficit of rainfall is reconstructed in the correct areas with the exception of the northwestern region. The small core of extreme deficit over Sagar (A) is recaptured. The small positive zone which appears in Assam, Andhra and Madras Province (B) and Sri Lanka is also present in the model. The narrow zone of positive departures at Lahore through Bikaner (C) is apparent, in good contrast to the adjacent negative zone.

Figure 3.3 shows the departure pattern calculated by regressing the 1921 transformed departures of the 30 oldest stations of the data set against the eigenvectors. The general appearance is similar to the results using 53 stations in the regression. The correlation between the two patterns (correlating all 53 stations, including the 23 not included in the 30-station regression) is + 0.87. A third regression using the 20 oldest stations produced the pattern shown in Figure 3.4 which has a correlation of + 0.84 (all 53 stations included in the correlation) with the 53-station model. The extreme appear to have too much emphasis, but the overall departure pattern is still acceptable.

Within the data set, the substitution model produces departure patterns that are quite similar when as few as 20 stations are used in the regression.

Section 3.3 - Model Performance Outside the Data Set for a Reduced
Number of Stations and Reduced Set of Base Eigenvectors

The year 1920 was chosen as the first test of the substitution model outside the primary data set (1921-1960) since the pattern of variability was expected to be similar to that analyzed in the original forty-year record. The eigenvectors from the forty-year record should be nearly identical to those which would have been obtained from an analysis of the forty-one year record including 1920. No comparison can be made between eigenvector coefficients as was done for 1921, but comparisons can be made between the observed and computed departure patterns when 52, 30, and 20 stations are used in the regression for the substitution model.

The normalized departure pattern of the transformed precipitation data for 1920 is shown in Figure 3.5. (The letter "M" appears in Figures 3.5, 3.12 and 3.16 to indicate stations not reporting July precipitation values for 1920, 1910, and 1899, respectively.) A small but intense zone of excess precipitation extends through Orissa, east, central and western Bihar and Northern Provinces with a maximum of 4.39σ (122.2 cm) at Daltonganj. Most of the remainder of the subcontinent experienced a deficit of rainfall, with minima at Silchar (-2.22σ equal to 25.0 cm), Vishakhapatnam (-2.59σ equal to 1.6 cm), Bangalore (-2.37σ equal to 2.3 cm) and Akola (-2.16σ equal to 6 cm).

Figure 3.6 shows the departure pattern derived from the substitute coefficients obtained by using 52 stations in the regression. Eigenvectors 3 and 4 were the dominant patterns. The departure pattern correlates well with the observed departures--the correlation coefficient is + 0.76. Although the extreme values are not reconstructed by the substitution model, the general trends are quite evident.

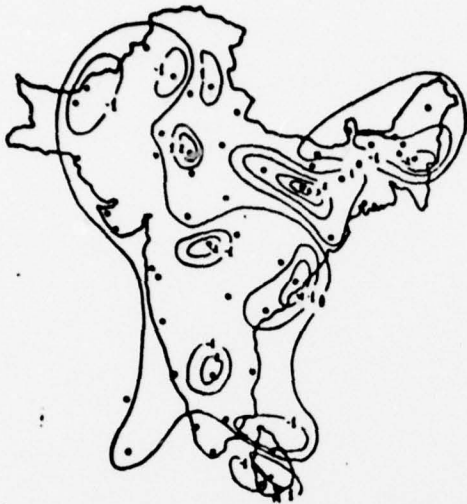


Fig. 3.5. Normalized departures for transformed values of measured precipitation for July 1920.

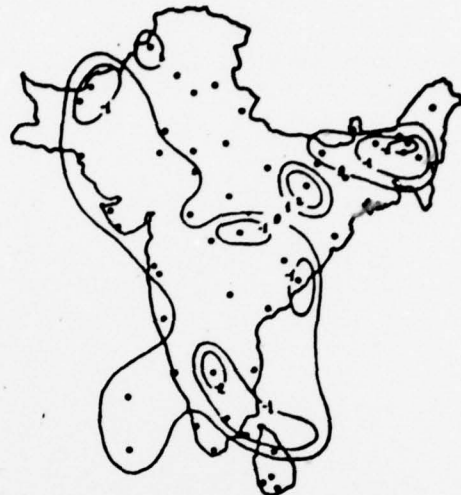


Fig. 3.6. July 1920 departures computed from the substitution model using 13 eigenvectors and 52 stations in the regression equation.

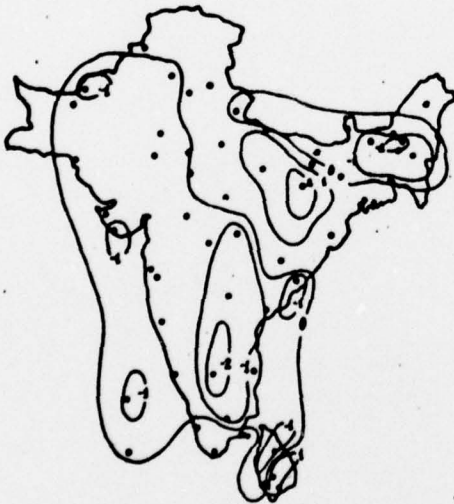


Fig. 3.7. July 1920 departures computed from the substitution model using 13 eigenvectors and 30 stations in the regression equation.



Fig. 3.8. July 1920 departures computed from the substitution model using 13 eigenvectors and 20 stations in the regression equation.

Figure 3.7 shows the departure pattern computed from the substitution model using 30 stations in the regression. The general trends established in the observed departure pattern are recreated in most sections, although the substitution model creates a zone of extreme values through the central peninsular region (A in Figure 3.7) that is not found in the observed departures. The correlation for the 30 stations included in the regression with the 52-station regression model was + 0.93. The correlation for the 22 stations not included in the regression was + 0.80. As expected, the correlation is slightly lower for the stations not included in the regression.

The departure pattern computed from the model using 20 stations in the regression is shown in Figure 3.8. Although the general pattern is correct, the model has begun to create serious discrepancies from the observed departure pattern. Sign reversals have occurred at Dumka (A in Figure 3.8) and Cuttack (B in Figure 3.8) that almost completely mask the positive departure that should be observed at the coast. Another sign reversal at Akola (C in Figure 3.8) creates a false positive zone in the midst of the observed negative departure. As in the 1921 case, several of the extreme values exceed the observed values, which is a reversal of the expected effect of eigenvectors to model the trends but reduce the amplitude of the variation. The correlation for the 20 stations included in the regression with the 52-station regression model was + 0.78. The correlation for the 32 stations not included in the regression was + 0.40. The use of the 20-station regression appears to create substitute records of substantially lower accuracy than do the 30- or 52-station regression models.

The sensitivity of the model to the number of eigenvectors used as the orthogonal base of the regression was tested by rerunning the 52, 30, and 20 station models using only the first four eigenvectors. As discussed in Section 2.4, these eigenvectors seem to represent the macroscale precipitation patterns without much interference of local effects.

Figure 3.9 shows the departure pattern for the four eigenvector-52 station regression model. The four eigenvector model creates a false positive zone (A in Figure 3.9) towards the northeast from the west coast while the thirteen eigenvector model does not. Additionally, the thirteen eigenvector model reconstructs the extrema more closely than the four eigenvector model. The correlation between the actual transformed departures and the four eigenvector-52 station model departures is .61, compared to the value of .76 for the thirteen eigenvector model.

Figure 3.10 shows the departure pattern for the four eigenvector-30 station regression model. This model is more like the actual departure pattern than the four eigenvector-52 station model, and it is not much different from the thirteen eigenvector-30 station model. The correlation for the 30 stations included in the regression with the four eigenvector-52 station regression was + 0.89. The correlation for the 22 stations not included in the regression was + 0.83.

Figure 3.11 shows the departure pattern for the four eigenvector-20 station model. Its replication of the actual transformed departure pattern appears to be much better than the thirteen eigenvector - 20 station model since the correct sign has been preserved at more stations and the extrema have not been exaggerated. The correlation for

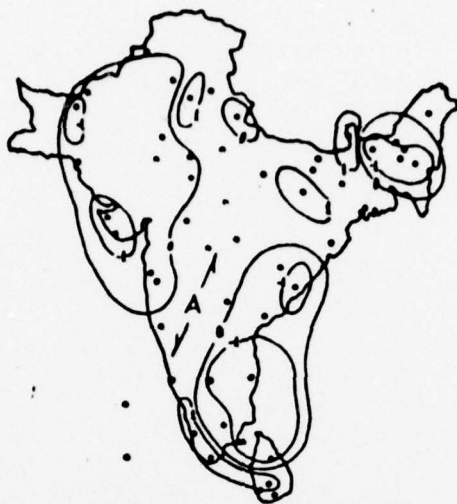


Fig. 3.9. July 1920 departures computed from the substitution model using 4 eigenvectors and 52 stations in the regression equation.



Fig. 3.10. July 1920 departures computed from the substitution model using 4 eigenvectors and 30 stations in the regression equation.

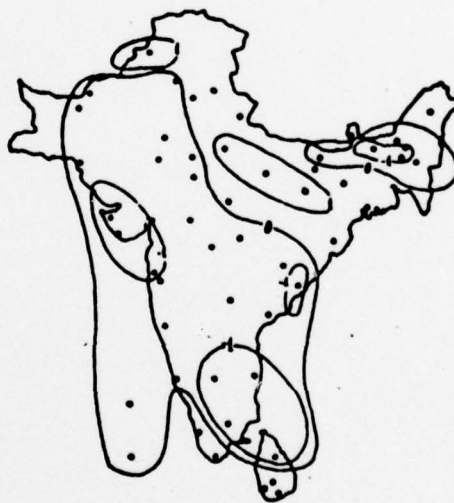


Fig. 3.11. July 1920 departures computed from the substitution model using 4 eigenvectors and 20 stations in the regression equation.

the 20 stations included in the regression with the four eigenvector-52 station model was + 0.74. The correlation for the 32 stations not included in the regression was + 0.84.

In Table 3.1 the results of both the four and thirteen eigenvector substitution model are contrasted to the actual measured precipitation for each of the twenty-two stations not used in the thirty-station regression equation.

Model departures were also calculated for 1910. Figure 3.12 shows the observed pattern with negative departures dominating the west coast and extending northward into West Pakistan, and positive departures apparent along the east coast extending into Punjab Province. There were still 52 stations reporting precipitation in 1910 from the original 53-station set. With all 52 stations used in the regression, the substitution model reconstructed a substantial amount of the original departure pattern. The substitution pattern is shown in Figure 3.13. The correlation between the patterns is .75. Eigenvectors 1 and 8 are the dominant patterns. The zones of greatest excess and deficit are well modelled, although the magnitudes were reduced, as expected. It appears that the eigenvector patterns are enough like the variance contained in the data for 1910 to allow a substantial reconstruction of the departure pattern.

The departure patterns for 1910 computed from the substitution model using 30 and 20 stations in the regression are shown in Figures 3.14 and 3.15. The 30-station model creates negative and positive areas similar to those of the 52-station model. However, there are a few major discrepancies in individual values. In contrast to this, the 20-station model has numerous discrepancies for individual stations

Table 3.1. Comparison of model output to actual precipitation (July 1920) for stations not used in the 30-station regression equation. The transformed normalized departures from the 30-station regression model have been converted to actual values by using the transformed mean and standard deviation of each station as listed in Table 2.1.

Number	Station	Measured precipitation (cm)	Model output	
			Thirteen eigenvector model (cm)	Four eigenvector model (cm)
4	Quetta	0.0	0.1	0.0
5	Kalat	0.2	0.3	0.0
8	Bikaner	14.2	7.1	1.7
15	Darbhangha	25.1	25.6	44.3
16	Dhubri	21.7	17.3	14.9
21	Cherrapunji	127.9	141.9	172.1
24	Dumka	56.9	28.2	45.0
28	Indore	50.9	24.0	24.1
48	Minicoy	13.4	17.5	18.0
50	Trincomalee	0.2	0.2	0.3
7	Mukteswar	33.4	28.8	25.1
11	Dibrugarh	43.7	59.4	68.3
13	Jodhpur	6.7	4.9	3.4
18	Kota	31.8	27.5	13.5
23	Daltonganj	122.0	67.0	57.0
27	Dwarka	8.6	7.1	1.0
31	Veraval	10.3	3.8	0.4
37	Begampet	14.9	7.7	12.9
44	Amini	18.1	3.2	15.4
45	Kodaikanal	8.2	4.4	6.7
47	Pamban	0.1	1.0	0.1
34	Jagdalspur	26.7	52.3	26.9
53	Hambantota	0.0	3.3	2.8

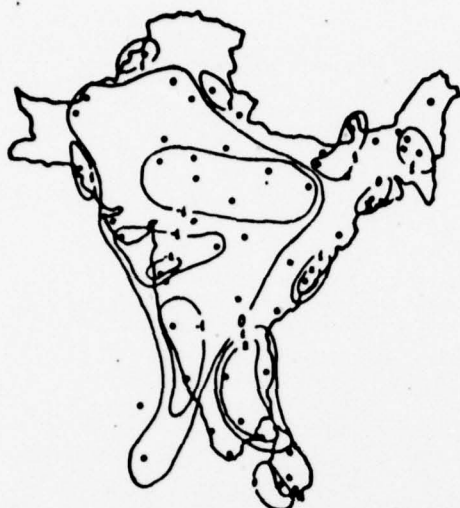


Fig. 3.12. Normalized departures of transformed values of measured precipitation for July 1910.

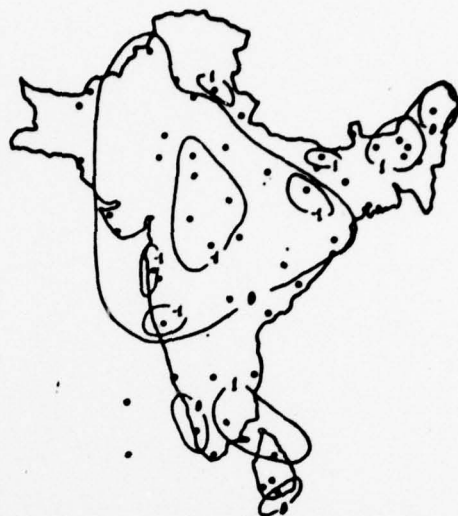


Fig. 3.13. July 1910 departures computed from the substitution model using 13 eigenvectors and 52 stations in the regression equation.



Fig. 3.14. July 1910 departures computed from the substitution model using 13 eigenvectors and 30 stations in the regression equation.



Fig. 3.15. July 1910 departures computed from the substitution model using 13 eigenvectors and 20 stations in the regression equation.

and several significant sign changes, especially along the southeastern coast (A in Figure 3.12) that diminish the value of the 13 eigenvector-20 station model for creating substitute records.

Analysis of the transformed departure patterns calculated from the model for 1920 and 1910 lead to several tentative conclusions concerning the substitution model. First, the model results do not degrade as a function of time, so the use of the 1921-1960 eigenvectors as an orthogonal base for the regression appears to be acceptable. Second, use of all thirteen eigenvectors provides the better estimate of the departure pattern when only a few stations are missing from the data set, while use of the first four eigenvectors produce better results when a significant function (22 or 32 out of 52) of the stations are missing from the data set. Third, the four eigenvector model (and to a lesser extent the thirteen eigenvector model) produces estimates that are nearly as accurate for the stations excluded from the regression as for those included in the regression equation.

Section 3.4 - Substitution Model for a Severe Drought Year

To check the model performance for a year with unusually large deviations, the departure pattern for 1899 was analyzed and reconstructed. As can be seen in Figure 3.16, July 1899 was a period of considerable drought through most of the subcontinent with the exception of a narrow zone of excess rainfall from Calcutta (A in Figure 3.13) to Mukteswar (B in Figure 3.13). With 47 stations reporting, seven reported negative departures exceeding -3σ .

The departure pattern calculated by using 47 stations on the regression is shown in Figure 3.17. Eigenvectors 3 and 2 were the dominant patterns. The overall anomaly pattern agrees very well with the observed departures; for the seven stations reporting departures more negative than -3σ , the model calculated values ranging from -1.8σ to -2.66σ . It appears that the model is able to reconstruct severe drought years with sufficient accuracy even twenty years outside the original data set.

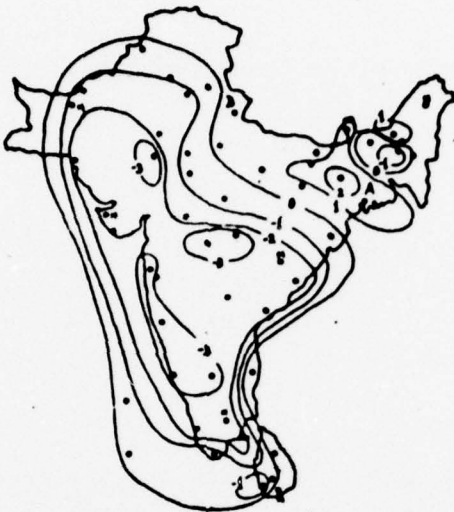


Fig. 3.16. Normalized departures of transformed values of measured precipitation for July 1899.

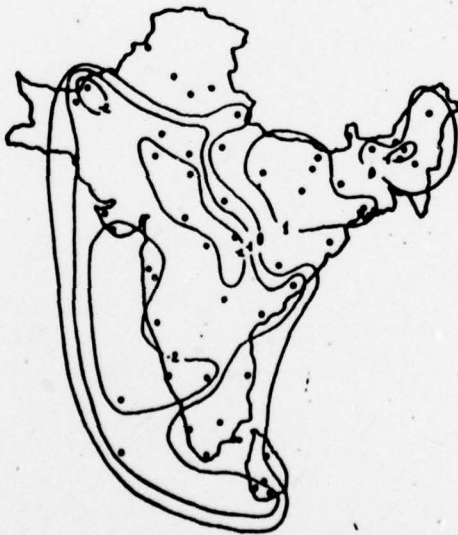


Fig. 3.17. July 1899 departures computed from the substitution model using 13 eigenvectors and 47 stations in the regression equation.

Chapter 4 - Conclusions

Section 4.1 - Results of this Study

The transformation of monsoon precipitation by the cube root function appears to provide the most nearly normalized data set, as compared to the original observations or the data transformed by the logarithmic function. Using normalized, cube root values of July precipitation, eigenvector analysis of 53 stations for the period 1921-1960 indicates that 13 out of a possible 40 eigenvectors account for 78% of the variance. The first four eigenvectors (accounting for 44% of the variance) best portray the macroscale precipitation of the July monsoon without interference of local effects.

A method of reconstructing an approximate precipitation record for stations missing data during years outside of the original data set is also discussed. Using either the first four or all thirteen eigenvectors as a linearly independent data base, the observed, normalized departures for stations which did report for the year in question are regressed against the eigenvector components of those stations in order to determine coefficients for each eigenvector for that year. These coefficients can then be used to generate a precipitation value for the missing stations, if eigenvector components are available for those stations from the original data analysis.

Substitute departure patterns of the transformed precipitation data were calculated for 1921, 1920, 1910, and 1899. Since the model results do not degrade as a function of time, the use of the 1921-1960 eigenvectors as an orthogonal base for the regression appears to be acceptable. Use of all thirteen eigenvectors provides the better estimate of the departure pattern when only a few stations are missing from the data

set, while use of the first four eigenvectors produce better results when a significant fraction (22 or 32 out of 52) of the stations are missing from the data set. The four eigenvector model (and to a lesser extent the thirteen eigenvector model) produces estimates that are nearly as accurate for the stations excluded from the regression as for those included in the regression equation.

Section 4.2 - Recommendations for Future Research

Additional studies are needed to determine if the eigenvector patterns remain constant over both time and space. The success of the substitution model for years outside of the original data set would suggest that this is a reasonable assumption, but more detailed comparisons should be made. In particular, the comparison between the two twenty-year periods of 1931-1950 and 1951-1970 would be especially interesting in view of the change in temperature trend which occurred around 1950. Also, the patterns from a series of eigenvector analyses based on increasing spatial resolution should be compared to determine if the eigenvector pattern changes significantly.

The highly variable monsoon precipitation of the Indian subcontinent was deliberately chosen as a test case for the substitution model. Since the model proved satisfactory in the test case, its accuracy should be checked in other geographic settings and for other meteorological parameters such as temperature and pressure. Also, the model should be verified on other even more variable months for the Indian subcontinent, such as the transition months of April and October or a northeast monsoon month such as January. To insure that individual storms during the July records did not overly bias the variance of the entire set, the total precipitation for July and August could also be analyzed, since this rough smoothing would eliminate the effect of many purely localized components.

Appendix I - Meteorological Applications of Variance Analysis

Orthogonal functions provide an efficient means of partitioning the total variance of a data set. The advantage of using orthogonal functions stems from the fact that the correlation between any two of the functions is exactly zero, so no orthogonal function can be a linear combination of any of the other orthogonal functions. Once the variance, expressed as a time series of normalized departures from the mean, has been characterized by a minimum number of orthogonal functions, the amount of variance explained by each function can be compared, and the relative importance of each function at different stations can be contrasted.

Sines and cosines have been extensively used as well as other orthogonal functions. The disadvantage of sine and cosine functions is that they must be harmonic multiples of each other. Time series with complex variance patterns can always be resolved into a series of periodic functions, but not as efficiently as if the functions need not be periodic.

If the data set is presented in the form of a variance-covariance matrix, a set of vectors can be derived which completely account for the variance of the matrix. These vectors, known as eigenvectors from the German word "eigen" meaning prime or characteristic, are essentially empirical polynomials bearing no harmonic resemblance to each other.

As an example of the physical basis of this eigenvector analysis, consider a data set consisting of ten years of observations of temperature at station A and station B. We can plot the normalized departures, computed by subtracting the mean from the actual observation

and dividing by the standard deviation for that station, for these two stations on a coordinate axis as shown in Figure A.1 where each point represents the values for a given year. This set of vectors is a physical representation of the variance of the data set, since the origin represents the mean value of the temperature at both A and B, and the square of the projection of each vector onto the two axes represents the variance at that station for that year. The length of the vector squared is proportional to the total variance associated with a given year. The greater the sum of the squared lengths of all the vectors, the greater variance there is to the total data set.

The objective of eigenvector analysis is to find a vector which can explain a maximum amount of the variance represented by the observations and then, once this variance has been removed, to find a second vector, perpendicular to its predecessor, which explains a maximum amount of the remaining variance. This process is continued until all variance is accounted for. The analytic technique can be described (following Kutzbach, 1967) as attempting to orient a "test" vector in such a way as to most resemble the observation vectors, which can be measured by calculating the sum of squares of the projections of the observation vectors onto the test vector. The larger the sum of squares, the more like the observation vectors the test vector has become. Since the test vector, when properly oriented, encompasses as much of the variance as is possible with a single vector, it is called the principle or primary eigenvector. In our example, note that several of the observation vectors fall into the first quadrant of the coordinate axes, so it is likely that the first eigenvector would also fall in this quadrant. Because of their orthogonality,

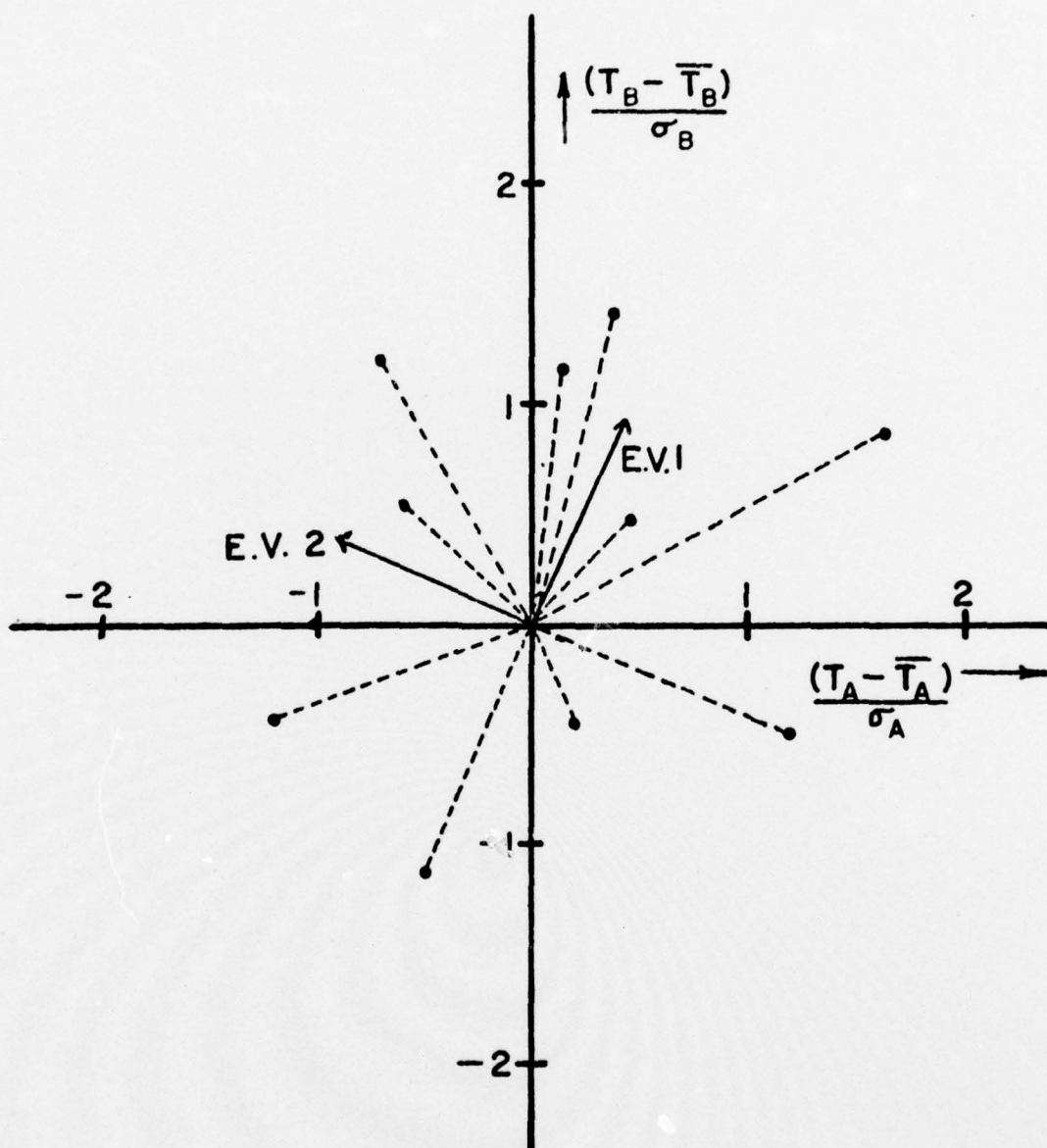


Figure A.1. Observations of normalized temperature departures at stations A and B. The approximate eigenvectors (E.V. 1 and E.V. 2) are also shown.

the second eigenvector must fall in the second or fourth quadrant.

Any of the observation vectors can be reconstructed exactly as a linear combination of these two eigenvectors. However, reconstruction of the entire variance of a data set is seldom needed in meteorology, since much of the variance may be due to localized effects, or noise rather than signal. Due to the maximization process just described, a small number of eigenvectors, often as few as 30 to 40% of the number of original observation vectors, will retain most of the large-scale variance of the data.

For an explanation of the matrix equations and procedures used in eigenvector analysis, the reader is referred to articles by Stidd (1967) and Kutzbach (1967). Briefly, the original data matrix is converted to a cross-product, covariance or correlation matrix from which a characteristic equation is derived whose roots form a set of characteristic or eigen-values. The sum of these eigenvalues equals the total variance of the converted data matrix. The unit vectors associated with each of the eigenvectors form a linearly independent set which defines a vector space of the same order as the converted data matrix.

The primary eigenvector is associated with the largest eigenvalue, and eigenvectors decrease in explained variance with the decreasing rank of the associated eigenvalue. The variance explained by each eigenvector equals its eigenvalue divided by the total variance of the data set (the sum of the eigenvalues).

Interpretation of eigenvectors normally consists of plotting the value of each eigenvector component associated with a given geographic location on a map of the area. The result will show areas of positive

and negative departures. The magnitude of the eigenvector component shows the relative contribution of any individual station to the variance explained by that eigenvector.

If the series of meteorological records contains a pattern of variance associated with a particular physical, climatic or dynamic phenomena (such as the windward or leeward effect of a mountain range on precipitation), that pattern of variance will usually dominate at least one of the eigenvector patterns. If the effect is known beforehand, an analyst can sometimes identify the pattern when it appears in an eigenvector. The yearly coefficients of the eigenvectors identify which eigenvector was most prevalent in a given year and might also help to identify which physical factors most influenced the variation of precipitation that year. Unfortunately, physical phenomena can seldom be clearly linked with an observed eigenvector pattern.

Within a set of eigenvectors, those with the larger eigenvalues normally display departure patterns characterized by shallow gradients and wide areal extent, while those with the small eigenvalues display patterns of alternating departures with steep gradients between them. Analogous patterns might be seen in a surface pressure map dominated by a single high pressure system, as compared to one in which several severe mesoscale storms are present. The eigenvectors with the larger eigenvalues explain variance within the data set that is found in all records of the set, while those with smaller values explain the variance "left over" once the macroscale variance is removed (local effects). If the individual records contain no widespread variance pattern, then all of the eigenvectors will display only localized effects.

Appendix II - Values of Components and Coefficients for Eigenvectors 1-22

The first five pages of this appendix list the 53 components for each of the first 22 eigenvectors obtained from the eigenvector analysis. The column index denotes the eigenvector while the row index denotes the component. Station identities for the row indices can be found in Table 2.1.

The final four pages, bearing the heading "C Transpose", list the coefficients of the eigenvectors for each year, starting with 1921 as year 1. The column index denotes the eigenvector while the row index denotes the year.

Graphical representations of the components and coefficients of the first thirteen eigenvectors are shown in Figures 2.6a through 2.6m.

References

- Blanford, H. L., 1889: Climates and Weather of India. MacMillan and Company, New York, 369 pp.
- Gangopadhyaya, M., P. Sreenivasan and R. Venkataraman, 1963: Some Characteristics of the Average Monsoon Rainfall Along the Coasts of India and Burma. Austral. Meteorol. Mag., 41, 23-41.
- Hirose, M. and J. Kutzbach, 1969: An Alternate Method for Eigenvector Computations. J. App. Met., 8, 701.
- Jagannathan, P., 1973: Trends and Periodicities in Rainfall over India. Monthly Weather Review, 101, 371-375.
- Jagannathan, P., and H. Bhalme, 1973: Changes in the Pattern of South-west Monsoon in India Associated with Sunspots. Monthly Weather Review, 101, 691-700.
- Jagannathan, P., and P. Rakhecha, 1972: Orthogonal Fields of Temperature Variation over Peninsular India. Indian J. Meteorol. and Geoph., 23, 317-326.
- Kutzbach, J., 1967: Empirical Eigenvectors of Sea Level Pressure, Surface Temperature and Precipitation Complexes over North America. J. App. Met., 6, 791-802.
- Lettau, K. and F. White, 1964: Fourier Analysis of India Rainfall. Indian J. Meteorol. and Geoph., 15, 27-38.
- Mooley, D. 1971: Independence of Monthly and Bimonthly Rainfall Over Southeast Asia During the Summer Monsoon Season. Monthly Weather Review, 99, 532-536.
- Mooley, D., 1973: Gamma Distribution Probability Model for Asian Summer Monsoon Monthly Rainfall. Monthly Weather Review, 101, 160-176.
- Normand, C. W. B., 1953: Monsoon Seasonal Forecasting. Quart. J. Roy. Meteorol. Soc., 79, 463-473.
- Ramage, C. S., 1971: Monsoon Meteorology. Academic Press, New York, 296 pp.
- Rao, P., 1965: Seasonal Forecasting - India. W. M. O. Technical Note 66, 17-30.
- Rao, P., 1971: Droughts in India. W. M. O. Technical Note 304.
- Stidd, C., 1967: The Use of Eigenvectors for Climatic Estimates. J. App. Met., 6, 255-264.

Subbaramayya, I., 1968: Interrelations of Monsoon Rainfall in Different Subdivisions of India. J. Meteorol. Soc. Japan, 46, 77-85.

Subrahmanyam, V., 1969: Some Aspects of Drought Climatology of the Dry Subhumid Zone of South India. J. Meteorol. Soc. Japan, 47, 237-244.

Walker, G. T., 1923: Correlations in Seasonal Variations of Weather, VIII. A Preliminary Study of World Weather. Mem. India Meteorol. Dept., 24, 75-131.