On the Prediction of India Monsoon Rainfall Anomalies

STEFAN HASTENRATH

Department of Meteorology, University of Wisconsin—Madison, WI 53706

(Manuscript received 25 August 1986, in final form 24 January 1987)

ABSTRACT

A complex of anomalies in the premonsoon large-scale circulation setting heralds the interannual variability of India summer monsoon rainfall. The most prominent precursors of precipitation anomalies are the latitude position of the upper-air ridge over India, apparently reflecting the persistence of the boreal winter wind regime and its consequence for the establishment of the summer upper-air circulation; the temperature in southern Asia and the adjacent North Indian Ocean waters, a factor instrumental in heat-low development and hence the establishment of meridional pressure gradients and lower-tropospheric airstreams from the Southern Hemisphere; and indices of the Southern Oscillation, capturing pressure departure patterns spanning the global tropics. Stepwise multiple regression is used to extract from this "anomaly complex" the variance most pertinent to the interannual variability of Southwest monsoon rainfall, observations of pertinent elements being available for the period 1939–81. Regression models developed on a portion of this record are then used to predict the summer monsoon rainfall anomalies of the years 1966–81.

The correlations between the various precursors and the rainfall anomalies vary in the course of 1939–81, being, on the whole, strongest in the 1950s and 1960s. While the April latitude position of the 500 mb ridge along 75°E proves to be the strongest predictor, performance is improved by inclusion of other elements representing preseason temperature and the Southern Oscillation. Correlation, root-mean-square error, bias, and absolute error are used as measures of forecast performance. A set of experiments with the dependent dataset, ending in 1965, indicates that a regression base period of about 20 yr is optimal for predictions into the independent portion of the record. Another set of experiments, in which the regression base periods are successively updated to the year immediately preceding the year to be forecast, shows no improvement of predictions over the fixed regression base periods. "Cross-validation" is not found less demanding than prediction proper. It is demonstrated that about half of the interannual variance in monsoon rainfall can be predicted from antecedent anomalies in the large-scale circulation setting.

1. Introduction

Climate prediction has been recognized as a task of high priority in both the World Climate Research Program (World Meteorological Organization, 1980, p. 42) and the U.S. National Climate Program (National Climate Program Office, NOAA, 1980, pp. 23–24). Results published in the 1980s (reviews in Hastenrath, 1985, p. 330–352; 1986) indicate that empirically based prediction of seasonal rainfall anomalies holds much promise for *certain* tropical regions. In India, attempts at long-range monsoon forecasting extend over a century. Investigations into the mechanisms of interannual climate variability have led to a considerable diagnostic understanding of monsoon rainfall anomalies in terms of the large-scale circulation (reviews in Cadet, 1985; Hastenrath, 1985, p. 283–294; Wu and Hastenrath, 1986), and have identified plausible precursors in the large-scale atmospheric and oceanic setting that may serve as input for the development of prediction methods. The present paper examines the potential of such predictors for the rainfall anomalies of the Indian summers southwest monsoon. In particular, complementing recent studies on the feasibility of empirically based climate prediction in low latitudes (Kung and Sharif, 1982; Nicholls, 1984; Hastenrath, 1987), three topics of general interest in tropical climate prediction are explored: secular variation of predictability, optimal length of dependent record, and updating of this input base.

2. Review of India climate prediction

A brief summary of work most pertinent for the present purposes must suffice here, while the two aforementioned recent reviews (Hastenrath, 1985, p. 330–352; 1986) contain a more extensive account and literature references.

More than a century ago, H. F. Blanford (Normand, 1953) issued monsoon forecasts, calling attention in particular to the negative relation between the winter snow in the Himalayas and the subsequent Southwest monsoon rainfall—a topic on which controversy still thrives in the professional journals of the 1980s (Hastenrath, 1985, p. 284). Seasonal monsoon forecasting was the motivation for Sir Gilbert Walker's (1910, 1924, 1928) pioneering work on the Southern Oscillation. Over the years, his forecasting schemes developed from four predictors in one formula for all of India to 28 predictors in six formulae differentiated regionally (Walker, 1910, 1924; Banerji, 1950). Walker left his mark on the long-range prediction activities of the India Meteorological Department to the present day. Walker (1928) stated the desired forecast performance as an 80% chance of success in a two-category scheme. On the part of the India Meteorological Department, the same percentage in a three-category scheme has been variously expressed as desired standard of performance (Jagannathan, 1960; Rao, 1965; Das, 1984, p. 33; 1986, p. 130). In appraising the predictive potential it is essential to distinguish between dependent and independent datasets, but it is not clear whether this distinction has always been made.

The evolution of the long-range forecasting endeavors of the India Meteorological Department in the course of the past half-century has been summarized in successive reviews (Banerji, 1950; Normand, 1953; Rao and Ramamoorthy, 1960; Jagannathan, 1960; Rao, 1965; Rao, 1976, p. 354–360; Das, 1984, p. 32–33; 1986, p. 126–136). At the time of Banerji's (1950) writing, the India Meteorological Department still used part of the predictors introduced by Walker, but no upper-air information. Rao and Ramamoorthy (1960) report that upper-air observations over India were added to the predictors familiar from Walker's work. The reviews by Jagannathan (1960) and Rao (1965) reflect the India Meteorological Department's progressive realization that the decaying wintertime upper-air wind regime in the premonsoon season may herald the quality of the subsequent Southwest monsoon. It is against this background of sustained work that we should appreciate Banerjee et al.'s (1978) study, who proposed as sole predictor of India's Southwest monsoon rainfall the latitude position of the 500 mb ridge along 75°E in April. Thapliyal (1982) used this parameter as sole input to an ARIMA (autoregressive integrated moving average) scheme to calculate the rainfall anomalies over peninsular India for four years, and claimed that his technique is superior to the multivariate method hitherto used by the India Meteorological Department. Mooley et al. (1986) chose the same parameter as sole predictor in a simple linear regression model to compute all-India monsoon rainfall anomalies for 4 yr. For the 1980s, Thapliyal (1982) reports that the India Meteorological Department's long-range forecasting method uses as predictors, temperature in India, South American pressure, and the April latitude position of the 500 mb ridge over India. From Walker's time to the 1980s, climate prediction in India seems to have evolved from a primarily statistical basis towards increased reliance on general circulation diagnostics. The potential of the pressure, wind, cloudiness, and sea surface temperature fields in the Indian Ocean as predictors of monsoon rainfall was explored by researchers outside India.

Based on diagnostic studies of the large-scale circulation, Kung and Sharif (1980, 1982) developed regression methods for forecasting both the Indian Southwest monsoon rainfall and the onset date over Kerala, South India. The input information consists of April upper-air patterns in the India–Australia region and sea surface temperature around India in the preseason. Kung and Sharif (1982) offer a measure of forecast performance by comparing the observed rainfall amounts with the values calculated from regression equations based successively on all years within the 1958–78 time span, except the particular year calculated. Expanding on extensive diagnostic studies in the Indian Ocean sector (Wu and Hastenrath, 1986), Wu (1985) identified plausible precursors of Southwest monsoon rainfall anomalies, including preseason temperature and pressure over land, temperature, pressure, wind, and cloudiness, in certain areas of the Arabian Sea, a Southern Oscillation index, and various upper-air parameters. These elements served as input to a stepwise multiple regression scheme. The aforementioned "cross-validation" (Efron, 1982) procedure of Kung and Sharif (1982) was used, whereby values of an annual rainfall index were calculated for individual years based on all other years of the 1951–70 period. The coefficient of correlation between the rainfall departures thus calculated and the observed anomalies is +0.92 (significant at the 0.1% level), a performance comparable to that of Kung and Sharif (1982). It can be expected, however, that the performance will be inferior for prediction proper—that is, for years beyond the end of the base period used in the construction of the regression models.

3. General circulation background

Nowhere else on earth is the seasonal reversal of the atmospheric and hydrospheric circulation as spectacular as in the Indian Ocean sector. Figure 1 highlights the salient features of the surface and upper-tropospheric circulation, while reference is made to Sadler (1975) and Hastenrath and Lamb (1979) for a detailed documentation of the annual cycle.

During the boreal winter (Fig. 1, part a), surface airstreams emanating from the interior of the Asian continent as Northeast monsoons sweep the Northern Indian Ocean, recurve to northwesterly near the equator, and meet the Southeast trades in a broad confluence zone situated in the Southern hemisphere. The upper-tropospheric circulation over southern Asia and the Northern Indian Ocean (Fig. 1, part c) is dominated by the Subtropical Westerly Jet located to the south of the Himalayas.

During the boreal summer (Fig. 1, part b), the surface circulation in the Indian Ocean sector is characterized by the Southeast trades of the Southern Hemisphere crossing the Equator and recurving to become the Southwest monsoon of the northern Indian Ocean and southern Asia, with a prominent speed maximum over the Arabian Sea. The upper troposphere to the south of the Himalayas is now dominated by strong easterlies [the tropical easterly jet (Fig. 1, part d)].

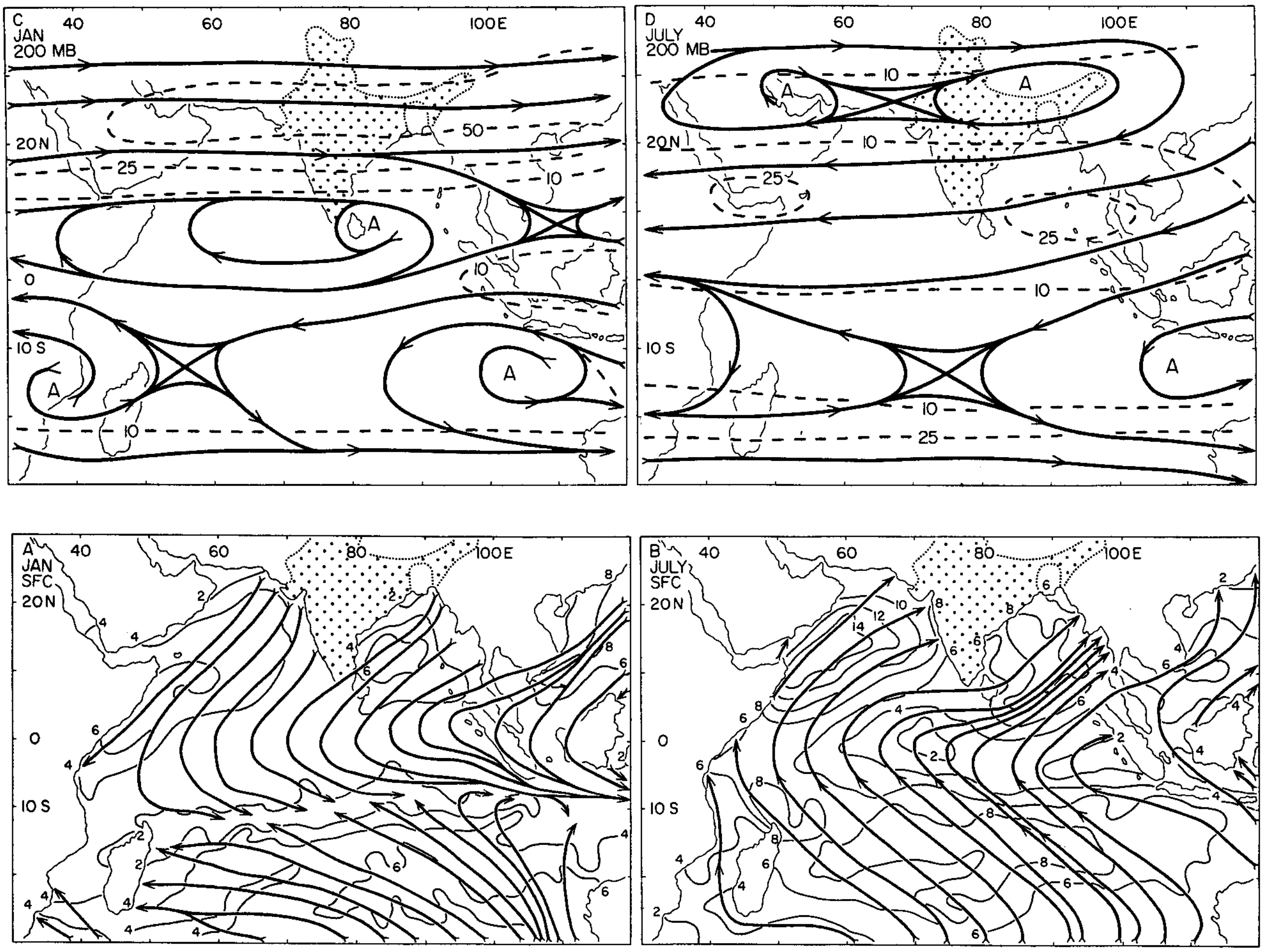

FIG. 1. Surface and 200 mb flow patterns in the Indian Ocean sector during (a, c) January and (b, d) July. Isotachs (broken lines) are in m s^{-1}, with different spacing for the surface (a, b) and 200 mb (c, d). India is shown by dot raster and dotted line (Ref. Fig. 2). (sources: Hastenrath and Lamb, 1979, charts 30 and 36; Atkinson, 1971, Figs. 4.10. and 4.11.).

The Southwest monsoon is the main rainy season for most of India. Accordingly, the changeover from the boreal winter to the summer circulation (ref. Fig. 1) is of particular concern for the prediction of monsoon rainfall anomalies. Indeed, extensive diagnostic analyses (Wu and Hastenrath, 1986; reviews in Hastenrath, 1985, p. 330–352; 1986) revealed particularly close associations between key atmospheric and oceanic indicators in the preseason and the quality of the subsequent Southwest monsoon. Preseason features deserving particular attention are summarized in the following, with consideration of the development of climate prediction methods as reviewed in section 2, and with direct reference to section 4. Surface temperature in South Asia and the adjacent North Indian Ocean waters (Fig. 2, BT, SST) relates to heat-low evolution, which in turn has a bearing on the cross-equatorial pressure gradient and monsoon airstreams in the lower troposphere. The persistence of winterly upper-air flow patterns—such as illustrated in Fig. 1, part c—may retard the evolution of the summertime wind regime exemplified by Fig. 1, part d, and thus herald anomalies in the boreal summer circulation (Fig. 2, L, D–N). Consistent with the review in section 2, it seems plausible that the precursors of an anomalous summer Southwest monsoon should be sought in the very tail of the boreal winter regime, rather than later on. Thus, after the onset of the Southwest monsoon, the temperature over land is strongly controlled by precipitation and the ensuing evaporative cooling, while the temperature of the Arabian Sea surface waters appears affected, in a variety of ways, by the wind field. Accordingly, land and ocean temperature departures and their correlations with the subsequent summer monsoon rainfall fade out from the preseason to the height of the Southwest monsoon (Wu and Hastenrath, 1986). Concerning the upper-air circulation patterns, the summer easterlies evolve only gradually after the decay of the distinct wintertime flow regime. Inasmuch as the Indian summer monsoon is intimately intertwined

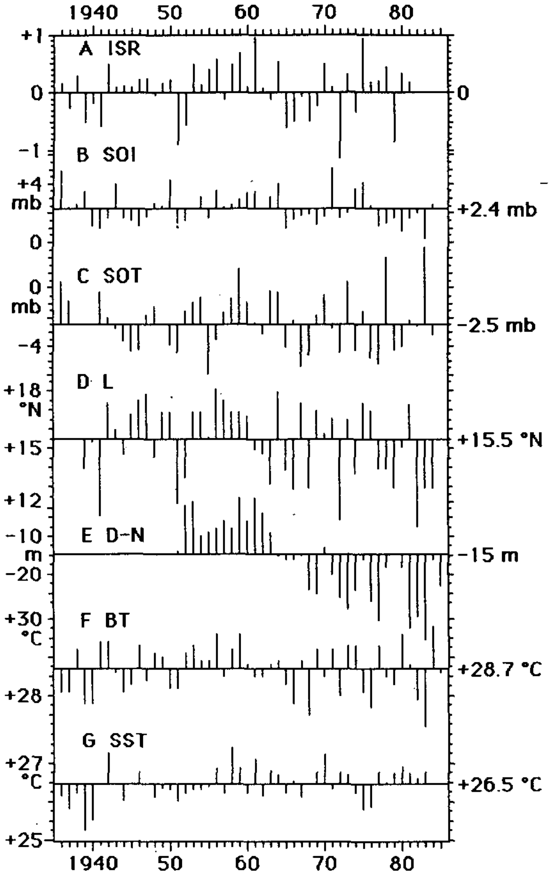

FIG. 2. Time series plots of (a) all-India index of summer monsoon rainfall ISR (source: Shukla, 1987); (b) Southern Oscillation index (pressure Tahiti minus Darwin; Parker, 1983) in April SOI; mean +2.4 mb. (c) Southern Oscillation tendency in March SOT; mean −2.5 mb; (d) latitude position (tenths of °N) of 500 mb ridge along 75°E in April *L*; mean 15.5°N (sources: Thapliyal, 1982; U.S. Weather Bureau, ESSA, NOAA, 1949–85); (e) difference of April 850 mb height (*m*), New Delhi minus Nagpur (D-N); mean −15 m, Refer to Fig. 3 for station locations. (source: U.S. Weather Bureau, ESSA, NOAA, 1949–85); (f) Bombay temperature (°C) in April BT, mean +28.7°C. Refer to Fig. 3 for station location. (sources: Smithsonian Institution, 1947; U.S. Weather Bureau, 1959; U.S. Weather Bureau, ESSA, NOAA, 1949–85); (g) sea surface temperature (°C) in Arabian Sea (5–20°N, 50–75°E) in January; mean +26.5°C; Series d to g are plotted centered on the reference period 1939–81, except 1951–81 for D-N.

with the circulation of the global tropics, a measure of the Southern Oscillation is of further interest for predictive purposes (Fig. 2, SOI, SOT).

4. Observations and methods

While the background for the present study includes extensive diagnostic investigations (Wu, 1985; Wu and Hastenrath, 1986) based on long-term ship observations and surface and upper-air station records in the Indian Ocean sector and the global tropics, only a limited number of time series are immediately used here, as illustrated in Figs. 2 and 3.

Southwest monsoon precipitation data for 31 subdivisions of India (Ref. Fig. 3) during 1901–81 were combined into an all-India summer rainfall index (ISR), by calculating for each year the arithmetic mean of the individual normalized rainfall departures. The spatial pattern of correlation between the all-India summer rainfall index and precipitation in the 31 subdivisions displayed in Fig. 3 shows largest values for the north-central portion of the country. A time series plot of the ISR for 1936–81 is presented in Fig. 2, part a. The use of a single bulk ISR for India as a whole appears desirable not only to compact the information and enhance data stability, but also because the primary purpose of this paper is to explore the overall potential of certain plausible general circulation precursors for the prediction of monsoon rainfall anomalies. Figure 3 illustrates that ISR represents best north-central India, and least the southern and eastern extremities of the country. It is conjectured that a regional differentiation of rainfall indices would allow one to enhance forecast performance, an issue to be pursued in later work.

As an index of the Southern Oscillation (SOI), the monthly pressure difference Tahiti minus Darwin (Parker, 1983) was chosen. Figure 2, part b, exhibits a 1936–85 time series plot of this index for April, the month during which SOI is most strongly correlated with ISR.

Monthly values of the Southern Oscillation tendency (SOT) were calculated as the difference between the

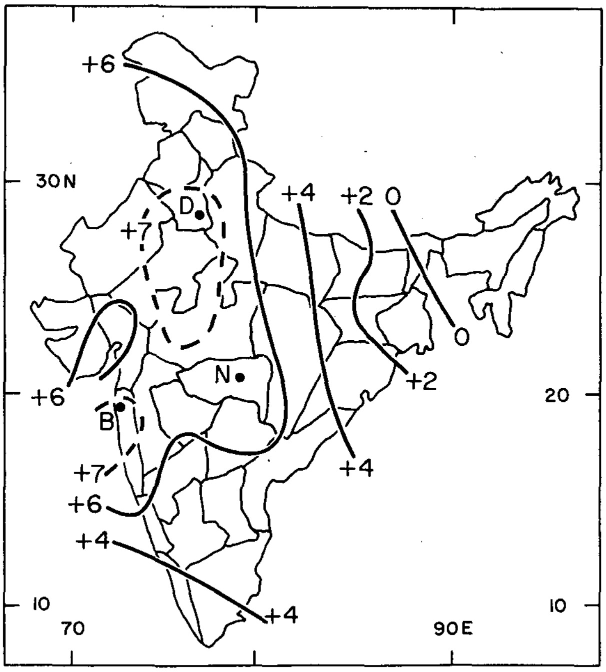

FIG. 3. Pattern of correlation (in tenths) between all-India index of summer monsoon rainfall and precipitation in the 31 subdivisions; period 1901–81. Dots and capital letters indicate the location of stations at New Delhi (D), Nagpur (N), and Bombay (B); Ref. Fig. 2.

SOI of the following minus that of the preceding month. A 1936–85 time series plot of SOT is reproduced in Fig. 2, part c, for March, the month for which SOT is most strongly correlated with ISR.

The April latitude position of the 500 mb ridge along 75°E (*L*) is plotted in Fig. 2, part d. The years 1939–81 were available from Thapliyal (1982), and the values for 1982–85 were produced from analysis of the upper-air data published by the U.S. Weather Bureau, ESSA, NOAA (1949–85). With reference to Fig. 3, it is noteworthy that Mooley et al. (1986) found *L* to be most highly correlated with monsoon rainfall in the north-central portion of India.

The difference of April 850 mb height New Delhi minus Nagpur (D–N) was recognized in earlier diagnostic work (Wu, 1985; Wu and Hastenrath, 1986) as a strong precursor of India monsoon rainfall anomalies. (Refer to Fig. 3 for the location of the upper-air stations.) Source data for this parameter are available in published form (U.S. Weather Bureau, ESSA, NOAA, 1949–85) for the years 1950–85 (except 1979). A time series plot is presented in Fig. 2, part e.

As a measure of preseason temperature over land, the April surface station temperature of Bombay (BT) was chosen, as this is regularly published (Smithsonian Institution, 1947; U.S. Weather Bureau, 1959; U.S. Weather Bureau, ESSA, NOAA, 1949–85). A time series of BT for the years 1936–85 is plotted in Fig. 2, part f.

Preseason sea surface temperature in the Arabian Sea (SST) was recognized in the earlier diagnostic work (Kung and Sharif, 1982; Wu, 1985; Wu and Hastenrath, 1986) as a precursor of monsoon rainfall anomalies. A time series compiled from long-term ship observations in the area 5–20°N, 50–75°E, for the month of January and years 1936–83 (except for gaps in 1941, 1943, 1945, and 1947) is plotted in Fig. 2, part g.

The time series plotted in Fig. 2, parts b and g, exhibit considerable similarity of interannual variations, both among themselves and with respect to the rainfall index ISR (Fig. 2, part a). These time series serve as input to a stepwise multiple regression scheme (Madison Academic Computing Center, 1980; Draper and Smith, 1981, p. 307–311; Hastenrath et al., 1984). The general regression equations are

$$\mathrm{ISR} = b_0 + b_1 X_1 + b_2 X_2 + \cdots b_n X_n \qquad (1)$$

where ISR is the predictand, the b_i the coefficients of the regression model, the X_i the regressors (variables plotted in Fig. 2, parts b to g), and n the number of regressors employed in the regression model. The model was used here in such a way that regression coefficients of variables retained can be accepted as differing from zero at the 5% significance level.

As a rule, a regression model is constructed on a "dependent" dataset. The formula in the form of Eq. (1), thus obtained, is then used in a predictive mode on an "independent" dataset. That is, the coefficients b_i are determined from a subset of the record, and these coefficients along with the observed values of X_i (elements plotted in Fig. 2, parts b to g) are then entered in Eq. (1) to calculate the values of ISR for a portion of the record not used in determining the coefficients b_i. Of the total 1939–81 record common to all series except D–N, the time span 1939–65 will be used to determine the regression coefficients, while the years 1966–81 are to be predicted. In light of the various previous analyses (ref. section 2) and the background work in this paper (Tables 1 and 2), the 1966–81 portion of the record is not claimed to be a "blind" sample. With this qualification, the 1939–65 portion of the record will be referred to in this paper, in brief, as a "dependent" dataset, and the portion of the record reserved for prediction experiments as an "independent" dataset.

The six elements plotted in Fig. 2, parts b to g, represent an ensemble of precursors of Southwest monsoon rainfall anomalies. They describe, in part, the atmosphere–hydrosphere general circulation setting; they should be thought of as pertaining to the overall "anomaly complex" of the combined tropical atmosphere–ocean–land system, and they are, indeed, correlated among each other (Ref. Table 1, parts a–c), or

TABLE 1. Matrix of correlation coefficients (in hundredths) between variables plotted in Fig. 2. Southern Oscillation index in April (SOI); Southern Oscillation tendency in March (SOT); April latitude position of 500 mg ridge (L); difference of April 850 mb height Delhi minus Nagpur (D–N); Bombay temperature in April (BT); January sea surface temperature in Arabian Sea (SST); Periods: A. 1939–65 (except 1951–81 for D–N); B. 1966–81; C. 1939–81 (except 1951–81 for D–N). Quenouille's (1952, p. 168) method was used to account for the reduction of the effective number of degrees of freedom due to persistence.

Variable	SOI	SOT	*L*	D–N	BT
		A. Period 1939–65			
SOT	+30				
L	+35	+05			
D–N	+17	+35	+15		
BT	+04	+42*	+26	+42	
SST	+18	+27	+44*	+22	+72**
		B. Period 1966–81			
SOT	−01				
L	+29	+21			
D–N	+15	+22	+03		
BT	0	+07	+33		
SST	−61	+21	−31	−11	+30
		C. Period 1939–81			
SOT	+17				
L	+34*	+15			
D–N	+25	+32	+15		
BT	+03	+26	+30	+18	
SST	+13	+22	+14	+05	+52**

* Significant at 5% level.

** Significant at 1% level.

TABLE 2. Correlation coefficients (in hundredths) between all-India index of summer rainfall ISR and variables plotted in Fig. 2, for various time intervals during 1939–81. Code is as follows: (SOI) Southern Oscillation index in April; (SOT) Southern Oscillation tendency in March; (*L*) April latitude position of 500 mb ridge; (D-N) difference of April 850 mb height Delhi minus Nagpur; (BT) Bombay temperature in April; (SST) January sea surface temperature in Arabian Sea; (SREG I) results of stepwise multiple regression model with SOI, SOT, *L*, D-N, BT, SST as input; (SREG II) model with SOI, SOT, *L*, BT as input; (VAR) variables retained by the stepwise multiple regression models. Quenouille's (1952, p. 168) method was used to account for the reduction of the effective number of degrees of freedom due to persistence.

	Time interval						
Variable	1941–50	1951–60	1961–70	1971–80	1939–65	1966–81	1939–81
SOI	+19	+63*	+72*	+19	+44	+20	+34
SOT	−40	+31	+46	+50	+17	+49	+33
L	+88**	+83**	+49	+80**	+72*	+70**	+72**
D-N	—	+40	+64	+44	+46	+22	+29
BT	−16	+78**	+40	+20	+42	+29	+37
SST	(+97)	+69*	+63	−44	+70*	−19	+34
SREG I	99	91	72	78	82	67	74
VAR	SST, BT	*L*, BT	SOI	*L*	*L*, SST	*L*	*L*, SST
SREG II	88	91	72	80	72	70	75
VAR	*L*	*L*, BT	SOI	*L*	*L*	*L*	*L*, SOT

* Significant at 5% level.
** Significant at 1% level.

share common variance. Correlations of these precursors with ISR and among each other differ between the dependent and the independent datasets (Table 1, parts a and b), which is of consequence for the forecast performance. Note that a large simple correlation between a certain element X_i and ISR does not necessarily make X_i the best predictor, because a large portion of its variance may be more effectively represented by some combination of other elements, each possessing a weaker simple correlation with ISR than X_i. In fact, the construction of a stepwise multiple regression model, Eq. (1), serves to extract from the "anomaly complex" the variance pertinent to the interannual variability of Southwest monsoon rainfall.

To appraise the forecast potential, four measures are used here: the correlation coefficient between the forecast ISR′ and the observed ISR, the root-mean-square error (RMSE), the bias (BIAS), and the absolute error (ABSE). The forecast period spans the 16 years 1966–81, and the aforementioned statistics are computed as follows (Nicholls, 1984):

$$\mathrm{RMSE} = [\sum_{66}^{81} (\mathrm{ISR}' - \mathrm{ISR})^2/16]^{0.5},$$

$$\mathrm{BIAS} = \sum_{66}^{81} (\mathrm{ISR}' - \mathrm{ISR})/16,$$

$$\mathrm{ABSE} = \sum_{66}^{81} |\mathrm{ISR}' - \mathrm{ISR}|/16,$$

where the summation extends over the 16 forecast years 1966–81. These measures are used to estimate the optimal length of the dependent record and the merits of updating the base period (sections 6 and 7).

Regarding the significance testing of correlation coefficients, it should be noted that geophysical time series are not, as a rule, serially independent. Therefore, Quenouille's (1952, p. 168) method was used to account for the reduction of the effective number of degrees of freedom due to persistence. This is based on the lag autocorrelations of the time series.

5. Secular variations

Major associations between the key monsoon parameters depicted in Fig. 2 persist throughout the 43 year record 1939–81, but Table 1 also shows that the interrelations between the various elements describing the premonsoon circulation setting differed between the 1939–65 and the 1966–81 intervals. More particularly, Table 2 examines the variations of relationships between the six premonsoon elements and the all-India rainfall index ISR in the course of the 43 year record.

For most elements, correlations are highest in the 1950s and 1960s (Table 2). This may be a factor for the very high performance obtained by Kung and Sharif (1982) for 1958–77 and Wu (1985) for 1951–70. Note that *L* exhibits consistently high correlations but with lowest values in the 1960s, in general agreement with the recent findings of Mooley et al. (1986). Comparing the 1939–65 and the 1966–81 portions of the record, *L* shows similar values for both; SOI, D-N, BT, and SST have higher correlations in the former interval, and SOT in the latter period. The stepwise multiple regression using all six precursors as input (SREG I) primarily yields somewhat higher correlations with ISR than any single element. The value of *L* is retained in most experiments, and as sole element in two cases. This regression exercise SREG I also shows that the two elements with incomplete record, D–N and SST, are of subordinate importance in describing the monsoon rainfall anomalies from the premonsoon setting. Therefore, these are excluded in the second exercise SREG II, for which the record is complete over the entire 1939–81 time span. The information content of *L* stands out most prominently. However, the description is improved by additional elements for the 1950s and the entire 1939–81 period, while for the 1960s *L* loses out to SOI and is not even retained. The bottom portion of Table 2, in particular, indicates some long-term variations in the relationships between the premonsoon circulation conditions and the rainfall anomalies of the subsequent summer monsoon; correlations being largest for the first two decades and smallest for the 1960s. In the remainder of this paper, analyses are confined to the elements used in the SREG II exercise—namely, SOI, SOT, *L*, BT, and ISR—for which the record is complete throughout 1939–81.

6. Optimal length of dependent record

The purpose of the present section is to estimate the most appropriate length of dependent record for the development of the regression model to be used in a predictive mode on an independent dataset—an issue considered before by Kung and Sharif (1982), Nicholls (1984), and Hastenrath (1987) for various low-latitude regions. On strictly statistical grounds, it may appear that the longest available dependent record would provide the best sampling, thus allowing the optimal description of relationships between precursors and the element to be predicted. In the meteorological perspective, however, it is realized that there are long-term changes in the large-scale circulation setting, which would entail variations in the relationships between the various precursors and ISR. Thus, an excessively long dependent record may contain much "outdated" information on the large-scale circulation setting, as compared to a shorter and more recent time span. These competing "statistical" and "meteorological" arguments are to be empirically pursued here.

In this section, the record up to 1965 is used as the "dependent" dataset, and the observations from 1966 onward as the "independent" dataset. Within the 1939–65 record, stepwise multiple regression models are constructed from base periods of varying length, but all ending with 1965. These regression models are then used to predict the summer monsoon rainfall anomalies in the independent dataset; that is, the years 1966–81. As measures of forecast performance the four statistics introduced in section 4 are used: namely, the correlation coefficient *R* between ISR′ and ISR, RMSE, BIAS, and ABSE. Results are summarized in Table 3, part A. Note that an experiment was also run for the 5-yr period 1961–65, but the scheme failed to produce a regression equation at the specified 5% significance level (ref. section 4).

Table 3, part A, shows the highest correlations between the forecast and the observed rainfall index for the longer regression base periods, but without deterioration from 27 yr to as short a base period as 20 yr.

TABLE 3. Appraisal of predictions of all-India summer monsoon rainfall index ISR during 1966–81. Quenouille's (1952, p. 168) method was used to account for the reduction of effective number of degrees of freedom due to persistence. *R*, Correlation coefficient between forecast ISR′ and observed ISR; root-mean-square error, RMSE = $[\Sigma\,(\mathrm{ISR}' - \mathrm{ISR}^2)/16]^{0.5}$; bias, BIAS = $\Sigma\,(\mathrm{ISR}' - \mathrm{ISR})/16$; absolute error, ABSE = $\Sigma\,|\mathrm{ISR}' - \mathrm{ISR}|/16$; all in hundredths; VAR: elements SOI, SOT, *L*, BT.

No. of years	Period regression	*R*	RMSE	BIAS	ABSE	VAR
			A. Fixed†			
27	1939–65	+70**	37	+1	32	*L*
25	41–65	+70**	38	+4	33	*L*
20	46–65	+70**	37	+3	33	*L*
15	51–65	+20	62	−0	53	SOI
10	56–65	+20	62	−1	52	SOI
			B. Updated#			
25	1941–65, etc.	+62*	41	−3	36	*L*, (SOI, SOT)
20	46–65, etc.	+53*	47	+1	40	*L*, (BT, SOI, SOT)
15	51–65, etc.	+35	58	+4	46	*L*, BT, SOI, SOT
10	56–65, etc.	+51*	49	+3	43	*L*, SOI, SOT

* Significant at 5% level.

** Significant at 1% level.

† Part A ("fixed"): Regression models, Eq. (1), used for forecasting the 16 years of the independent data set were constructed from base periods of varying length as indicated in left portion of table, but all ending in 1965 (i.e., the 5 yr regression model is constructed from the years 1961–65, . . . the 27 year regression model from the years 1939–65).

Part B ("updated"): Forecasts are based on regression models with dependent regression base period always ending with the year preceding the predicted year. (For example: in the case of the 15 year regression base period the dependent years 1951–65 are used to predict 1966, . . . the dependent years 1957–71 to predict 1972, etc.)

At still shorter intervals the correlations drop off drastically, and, in fact, no model could be constructed at 5 yr. Similarly, RMSE and ABSE remain virtually invariant from 27 to 20 yr, but increase drastically for the shorter periods, while BIAS is small at all lengths of base periods. Note that for the intervals of 20 yr and larger, the April latitude position of the 500 mb ridge is retained as sole predictor and shows the best performance, whereas for the shorter periods SOI prevails. This appears, in part, plausible from Table 2, which identifies for the SREG I and SREG II experiments SOI retained as sole regressor. The four measures of forecast performance presented in Table 3 are internally consistent in suggesting about 20 years as the most meaningful length of the regression base period but not substantially longer. This is broadly consistent with the findings of Kung and Sharif (1982). A scatter plot for the 20 yr experiment is reproduced in Fig. 4. The highly significant correlation between the forecast and the observed rainfall index indicates that about half of the interannual variability of Southwest monsoon rainfall can be predicted from antecedent circulation departures.

7. Updating of base period

Long-term changes in the large-scale circulation setting have the consequence that the most recent information of the dependent record may be most valuable for predictions into the independent time interval. Accordingly, the present section explores the usefulness of updating the regression model base period. Regression models in the form of Eq. (1) were constructed for all of the various lengths of base periods listed in Table 3, part A. However, in addition to the models with base periods of the dependent dataset all ending with 1965, models were also constructed with base periods ending in 1966, 1967, . . . 1980. That is, for each of the four record lengths (i.e., 10, 15, 20, 25 yr) listed in Table 3, part B, a total of 16 regression models were constructed. For the 5 yr base period the scheme was capable of producing regression models for some, but not all, of the 16 combinations considered here. Therefore, particulars are not detailed here. Each of the models was then used to predict the rainfall for the year immediately following the last year of the respective regression base period. Results are summarized in Table 3, part B, and also illustrated in Figs. 4 and 5.

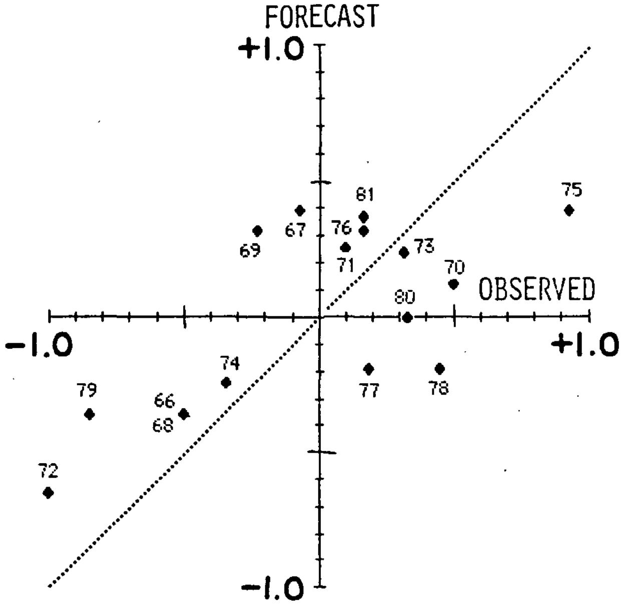

FIG. 4. Scatter diagram of forecast versus observed all-India summer monsoon rainfall index. Regression base period is 1946–65 and forecast period 1966–81. Numbers indicate the years and dotted line 45° angle. Correlation coefficient $r = +0.70$ is significant at the 1% level. Quenouille's (1952, p. 168) method was used to account for the reduction of the effective number of degrees of freedom due to persistence; RMSE = 37, BIAS = +3, ABSE = 33.

Comparison of Table 3, part B, with part A bears out the overall remarkably inferior performance of the "updated" models: correlations are smaller, RMSE and ABSE larger, while BIAS remains as small as for the "fixed" models. Note that L remains the most important predictor for the 25- and 20-year models, although the other three elements are also used in some years. By comparison, in the 15 and 10 yr models L does not serve as predictor for all years. Even for these intervals, L is most important for the most recent portion of the 1966–81 prediction period, broadly consistent with the SREG I and SREG II results for the 1971–80 decade presented in Table 2. For the 20-yr regression base period, the updated version (Fig. 5), illustrates a much larger scatter than the fixed version, Fig. 4. In fact, the updated model proves capable of explaining only about one-fourth of the interannual rainfall variance, as opposed to nearly one-half for the fixed version. Table 3 and Figs. 4 and 5 thus indicate that, at least with the regressors used and the time intervals examined here, "updating" is inferior to the use of "fixed" regression models. It is conjectured that this unexpected behavior may be due to secular variations in the large-scale circulation setting and predictability. In this context it should be noted that in a case study of the predictability of Java monsoon rainfall (Hastenrath, 1987) updating likewise failed to improve forecast performance.

8. A "cross-validation" experiment

In their "cross-validation" experiments over two decades summarized in section 2, both Kung and Sharif (1982) and Wu (1985) achieved a highly significant correlation between calculated and observed rainfall anomalies. It was anticipated (Hastenrath, 1985, p. 339–352; 1986) that the method would yield a substantially inferior performance for predictions proper—that is, for years beyond the last year on which the regression model is based. This contention could not

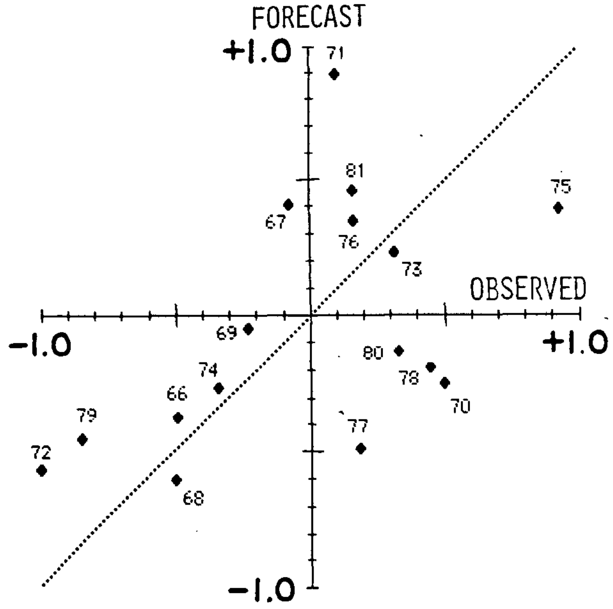

FIG. 5. Scatter diagram of forecast versus observed all-India summer monsoon rainfall index for the years 1966–81. The length of the regression base period is always 20 years, but forecasts are made from models with regression base periods ending 1 yr before the predicted year (example: the dependent record 1946–65 serves to predict 1966, . . . , the record 1953–72 to predict 1973, etc.). Numbers indicate the years and dotted line 45° angle. Correlation coefficient $r = +0.53$ is significant at the 5% level. Quenouille's (1952, p. 168) method was used to account for the reduction of the effective number of degrees of freedom due to persistence. RMSE = 47, BIAS = +1, ABSE = 40.

be verified from the short database used in Kung and Sharif (1982) and Wu (1985). Therefore, it was found desirable to explore this issue here.

Twenty regression models were always constructed from 19 of the 20 years in the period 1962–81. These 20 models were then used to calculate the monsoon rainfall anomaly for the 1-yr not included in the development of the respective regression model. A base period of about two decades was chosen in light of the results of section 6, and for consistency with the experiments of Kung and Sharif (1982) and Wu (1985). Results are plotted in Fig. 6. At most, two regressors were retained (namely, L in all 20 and SOT in 19 of the regression models). The correlations between calculated and observed values for both the entire 1962–81 record and the 1966–81 interval used in sections 6 and 7 are considerably larger than for the "updated" (section 7, Fig. 5), but smaller than for the "fixed" (section 6, Fig. 4) experiments. Thus the results do not support the conjecture that "cross-validation" is less demanding than prediction proper. The very high performance obtained in the cross-validation experiments of Kung and Sharif (1982) for 1958–77 and Wu (1985) for 1951–70 appears largely due to the circumstance that these intervals include a favorable phase in the secular variation of predictability (Ref. Table 2).

9. Conclusions

Extensive investigations in India and abroad have led to a general understanding of the interannual variability of monsoon precipitation and have identified certain preseason departure characteristics of the combined atmosphere–ocean–land system that herald monsoon rainfall anomalies. In brief, these include three categories of parameters indicative of (i) upper-air flow over India, (ii) temperatures in South Asia and the adjacent waters, and (iii) the Southern Oscillation.

The upper-air flow conditions in the premonsoon season reflect the interannually varying persistence of the boreal winter wind regime, which must be regarded as consequential for the development of the summer flow patterns. The most prominent known precursor of this kind is the April latitude position of the 500 mb ridge along 75°E. A much less powerful parameter pertaining to this category is the difference of the April 850 mb height in Delhi minus Nagpur. The preseason temperature in the northern portion of the Indian Ocean sector relates to heat-low development, which

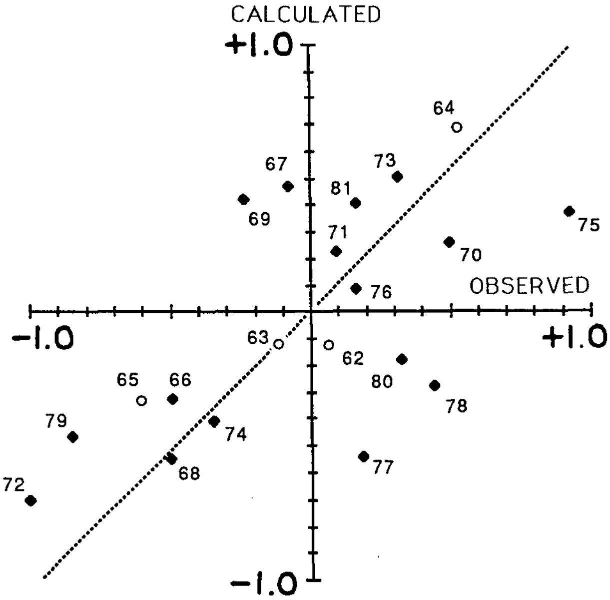

FIG. 6. Scatter diagram of calculated versus observed all-India summer monsoon rainfall index for the years 1962–81 using "jackknifing". Values for individual years are computed from regression models constructed successively from all of the years 1962–81 except the particular year to be calculated; i.e., the value for 1962 is calculated from a regression model based on the years 1963–81, . . . the value for 1968 from the years 1962–67 and 1969–81, . . . etc. Numbers indicate the years and dotted line 45° angle. The years 1962–65 are entered as open circles, all other years as solid dots. The correlation coefficient between calculated and observed values is +0.65 for the entire 1962–81 period, and +0.61 for the 1966–81 interval, significant at the 1 and 5% levels, respectively. Quenouille's (1952, p. 168–170) method was used to account for the reduction of the effective number of degrees of freedom due to persistence. 1962–81: RMSE = 39, BIAS = +0, ABSE = 32. 1966–81: RMSE = 42, BIAS = +2, ABSE = 36.

in turn is relevant to the establishment of a northward pressure gradient and cross-equatorial, lower-tropospheric flow from the Southern Hemisphere. The Bombay temperature in April was readily available for the present study, but other indicators belonging to this category have been widely used since Walker's time. (Walker was likewise aware of the associations between the vagaries of the Southwest monsoon and pressure seesaws of the Southern Oscillation type spanning the global tropics.) The surface pressure difference Tahiti minus Darwin was adopted here as a convenient measure of the Southern Oscillation. However, again, other parameters such as the "South American pressure" referred to in section 2 appear to pertain to the same category.

Precursors are interrelated not only within, but also between the aforementioned three broad categories. Stepwise multiple regression was used to extract from the atmosphere–ocean–land "anomaly complex" the variance most pertinent to the interannual variability of summer monsoon rainfall. In addition to the correlation coefficient, three other statistics were used to appraise the forecast performance, namely root-mean-square error, bias, and absolute error. Particular attention was given to secular variations, optimal length of dependent record, and updating of the input base.

Long-term variations are apparent in the correlations between the various precursors and monsoon rainfall anomalies, with, on the whole, the closest relationships in the 1950s and 1960s. The April latitude position of the 500 mb ridge along 75°E is the strongest predictor, but performance is improved by inclusion of Bombay April temperature and certain indices of the Southern Oscillation. Predictions were made into the "independent" portion of the record from regression models constructed on the basis of a "dependent" dataset. For predictions with a fixed "dependent" base period of the regression model it is found that a regression base period of about 20 yr yields the optimal performance. Successive updating of the regression base period to the year immediately preceding the year to be predicted does not improve the forecast performance. The limited experiments conducted here do not support the contention that "cross-validation" is less demanding than prediction proper. This study confirms that about half of the interannual variance of an all-India summer monsoon rainfall index can be predicted from antecedent anomalies in the large-scale circulation. It is conjectured that availability of more representative elements to describe the preseason temperature conditions and the Southern Oscillation, along with a regional differentiation of the predictand, may further improve the forecast performance.

Acknowledgments. This study was supported by U.S. National Science Foundation Grant ATM-8413575. Stephan Jansen did the computer programming. I thank P. K. Das, Ernest Kung, Peter Lamb, and anonymous reviewers for comments on a draft version of this paper.

REFERENCES

Atkinson, G. D., 1971: Forecaster's guide to tropical meteorology. Technical Report No. 240, USAF, Air Weather Service, 341 pp.

Banerjee, A. K., P. N. Sen and C. V. R. Raman, 1978: On foreshadowing Southwest monsoon rainfall over India with midtropospheric circulation anomaly of April. *Indian J. Meteor. Hydrol. Geophys.,* **29,** 425–431.

Banerji, S. K., 1950: Methods of foreshadowing monsoon and winter rainfall in India. *Indian J. Meteor. Geophys.,* **1,** 4–14.

Cadet, D. L., 1985: The Southern Oscillation over the Indian Ocean. *J. Climatol.,* **5,** 189–212.

Das, P. K., 1984: The monsoons—a perspective. India National Science Academy, Perspective Report Series 4, New Delhi, 52 pp.

——, 1986: Monsoons. *Fifth IMO Lecture,* Geneva, World Meteorological Organization, WMO-No. 613, 155 pp.

Draper, N. R., and H. Smith, 1981: *Applied regression analysis.* Wiley, 709 pp.

Efron, B., 1982: *The Jackknife, the Bootstrap, and other Resampling Plans.* Monograph No. 38, Society for Industrial and Applied Mathematics 92 pp.

Hastenrath, S., 1985: *Climate and circulation of the tropics.* Reidel, 455 pp.

——, 1986: On climate prediction in the tropics. *Bull. Amer. Meteor. Soc.,* **67,** 692–702.

——, 1987: Predictability of Java monsoon rainfall anomalies: a case study. *J. Climate Appl. Meteor.,* **26,** 133–144.

——, and P. J. Lamb, 1979: *Climatic atlas of the Indian Ocean Part 1. Surface climate and atmospheric circulation. Part 2. The oceanic heat budget,* University of Wisconsin Press, 114 and 110 pp.

——, M.-C. Wu and P.-S. Chu, 1984: Towards the monitoring and prediction of Northeast Brazil droughts. *Quart. J. Roy. Meteor. Soc.,* **110,** 411–425.

Jagannathan, P., 1960: *Seasonal forecasting in India, a review.* Meteorological Office, Poona, 79 pp.

Kung, E. C., and T. A. Sharif, 1980: Regression forecasting of the onset of the India summer monsoon with antecedent upper air conditions. *J. Appl. Meteor.,* **19,** 370–380.

——, and ——, 1982: Long-range forecasting of the Indian summer monsoon onset and rainfall with upper-air parameters and sea surface temperature. *J. Meteor. Soc. Japan,* **60,** 672–681.

Madison Academic Computing Center, 1980: STATJOB summary, stepreg. University of Wisconsin, Madison, 50 pp.

Mooley, D. A., B. Parthasarathy and G. B. Pant, 1986: Relationship between Indian summer monsoon rainfall and location of the ridge at the 500 mb level along 75°E. *J. Climate Appl. Meteor.,* **25,** 633–640.

National Climate Program Office, NOAA, 1980: National Climate Program. Washington, D.C., 101 pp.

Nicholls, N., 1984: The stability of empirical long-range forecast techniques: a case study. *J. Climate Appl. Meteor.,* **23,** 143–147.

Normand, C. W. B., 1953: Monsoon seasonal forecasting. *Quart. J. Roy. Meteor. Soc.,* **79,** 463–473.

Parker, D. E., 1983: Documentation of a Southern Oscillation index. *Meteor. Mag.,* **112,** 184–188.

Quenouille, M. A., 1952: *Associated measurements.* Butterworths, 242 pp.

Rao, K. N., 1965: *Seasonal forecasting—India.* WMO Technical Note No. 66, WMO No. 162, TP 79, pp. 17–30.

Rao, R. N., and K. S. Ramamoorthy, 1960: Seasonal (monsoon) rainfall forecasting in India. *Symposium on Monsoons of the Word,* R. R. Ramanathan, and U. K. Bose, Eds., Hindi Union Press, 271 pp.

Rao, Y. P., 1976: *Southwest monsoon. Synoptic Meteorology, Meteorological Monograph,* India Meteorological Department, No. 1/1976, 307 pp.

Sadler, J. C., 1975: The upper tropospheric circulation over the global tropics. Department of Meteorology, University of Hawaii, UHMET-75-05, 33 pp.

Shukla, J., 1987: Interannual variability of monsoons. in *Monsoons.* Fein, J., and P. Stephens, Eds., Wiley Interscience Publishers, (in press).

Smithsonian Institution, 1947: World weather records, 1931–40. Smithsonian Miscellaneous Collections, 105, 646 pp.

Thapliyal, V., 1982: Stochastic dynamic model for long-range prediction of monsoon rainfall in peninsular India. *Mausam,* **33,** 399–404.

U.S. Weather Bureau, 1959: World weather records, 1941–50, Washington, D.C., 1361 pp.

U.S. Weather Bureau, ESSA, NOAA, 1949–85: Monthly climatic data for the World. Asheville, N.C.

Walker, G. T., 1910: Correlation in seasonal variations of weather II. *Mem. Indian Meteor. Dept.,* **21,** 22–45.

——, 1924: Correlation in seasonal variations of weather III: the local distribution of monsoon rainfall. *Mem. Indian Meteor. Dept.,* **23,** 23–40.

——, 1928: Ceará (Brazil) famines and the general air movement. *Beitr. Phys. Freien Atmos.,* **14,** 88–93.

World Meteorological Organization, 1980: Outline plan and basis for the World Climate Programme 1980–83. WMO-No. 540, Geneva, 64 pp.

Wu, M.-C., 1985: On the prediction of the Indian summer monsoon. Papers in Meteorological Research (Taipeh), **8**(1), 34–44.

——, and S. Hastenrath, 1986: On the interannual variability of the Indian monsoon and the Southern Oscillation. *Theor. Appl. Climatol.,* **36,** 239–261.