Regression Forecasting of the Onset of the Indian Summer Monsoon with Antecedent Upper Air Conditions

Author(s): Ernest C. Kung and Taher A. Sharif

Source: *Journal of Applied Meteorology (1962-1982)*, April 1980, Vol. 19, No. 4 (April 1980), pp. 370-380

Published by: American Meteorological Society

Stable URL: https://www.jstor.org/stable/26179098

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

is collaborating with JSTOR to digitize, preserve and extend access to *Journal of Applied Meteorology (1962-1982)*

JSTOR

This content downloaded from 130.190.247.198 on Thu, 23 Feb 2023 11:10:21 UTC
All use subject to https://about.jstor.org/terms

Regression Forecasting of the Onset of the Indian Summer Monsoon with Antecedent Upper Air Conditions[1,2]

ERNEST C. KUNG AND TAHER A. SHARIF

Department of Atmospheric Science, University of Missouri, Columbia 65211

(Manuscript received 5 November 1979, in final form 12 December 1979)

ABSTRACT

It is shown that the recorded onset dates of the summer monsoon in southwestern India can be closely related functionally to the antecedent upper air conditions. The antecedent upper air conditions are represented by April mean values of the daily upper air parameters at 100 and 700 mb from 1958 to 1978. It is further demonstrated that the multi-regression scheme on the basis of such a functional relationship may be utilized in an objective forecast of the onset date. The forecasted dates are shown to be reasonably close to the recorded onset dates. Various aspects of the approach and scheme are discussed. Because of the time variations of the patterns of the general circulation, a constant updating of the regression equation should be an integral part of the multi-regression forecasting scheme.

1. Introduction

As one of the most significant regional patterns of the summer circulation in the Northern Hemisphere resulting from land-sea contrasts and orographic features, the Asian summer monsoon presents interesting problems in the study of the general circulation and in the development of a forecasting scheme. The characteristic large-scale fields of circulation in association with the monsoon have been studied extensively (e.g., Ananthakrishnan, 1977; Findlater, 1969; Flohn, 1968; Krishnamurti, 1971; Lockwood, 1974; Manabe *et al.*, 1974; Murakami, 1976; Ramage, 1971; Ramage and Raman, 1970; Shukla, 1975; Sikka, 1977; Singh *et al.*, 1978; Staff Members of the Academia Sinica, 1958; Walker, 1923; Winston and Krueger, 1977; Yeh and Chang, 1974; Yoshino, 1971). These studies indicate that among the large-scale features associated with the monsoon, three of them may stand out for their fundamental importance in the onset and maintenance of the monsoon: 1) the large warm-core anticyclone formed in the upper troposphere and lower stratosphere over the Tibetan Plateau in response to the heating by the plateau; 2) the intense cross-equatorial southwesterly flow over the Indian Ocean which accounts for the significant portion of moisture transport to the monsoon rainbelt over India; and 3) the pronounced temperature contrast between the land and water masses which provides a necessary large-scale vertical component of the motion in forming the so-called monsoon meridional circulation system.

Despite the extensive studies already in existence, the complex physical mechanisms of the monsoon system are yet to be systematically investigated in various scales of dynamics and energetics to provide the long-range forecasting ability concerning the monsoon. The comprehensive 1979 field observation periods of the GARP Monsoon Experiment (MONEX) and the MONEX related researches are one significant step in the organized scientific efforts in this direction. On the other hand, before the eventual goal of the long-range forecasting capability of the mathematical-dynamical models is achieved through advanced knowledge of the physical system, the development of an empirical forecasting scheme is a desirable step in view of the economic importance of the Asian summer monsoon. Lorenz (1956) gave a meticulous discussion on the dynamical basis for statistical weather prediction, and stated that the dynamic equations themselves justify attempts to predict the future state of the atmosphere by empirical methods even when the data are restricted in extent and in kind. He particularly pointed out that the standard methods of linear prediction may be used to good advantage with the data of the near past to the extent that the information concerning the nonlinear processes is implicitly contained in the near past behavior of the atmosphere. The concept presented by Lorenz seems to be applicable in the case of formulating an empirical forecasting scheme of the monsoon. Namias (e.g.,

[1] Research supported by the Division of Atmospheric Sciences, Global Atmospheric Research Programs, National Science Foundation.

[2] Contribution from the Missouri Agricultural Experiment Station, Journal Series No. 8351.

0021-8952/80/040370-11$06.75
© 1980 American Meteorological Society

This content downloaded from 130.190.247.198 on Thu, 23 Feb 2023 11:10:21 UTC
All use subject to https://about.jstor.org/terms

1959, 1973, 1976, 1978) extensively studied the empirical long-range forecasting schemes on the basis of short-period climatic fluctuations of multiple causes and nonlinear feedback loops in various locations in the system of the general circulation. The series of work by Namias well demonstrates the feasibility of the empirical approach in long-range forecasting.

It is well known that the period from March to May, recognized as the hot weather season in India, is associated with the distinct transition in values of various meteorological parameters, and is the prelude to the onset of the southwest Indian monsoon. Ananthakrishnan (1970, 1977) in particular demonstrated that these patterns of transition are most visible in the upper and lower troposphere. As will be shown in this paper the interannual variations of the meteorological parameters during the premonsoon period also appear significant. It is the purpose of this paper to demonstrate that a functional relationship may be obtained between the observed date of the monsoon onset and certain parameters which describe the mean upper air condition during a period preceding the onset. An attempt will then be made to utilize such a regression function in an objective prediction of the monsoon onset date, and the results of the attempt will be examined in various aspects.

2. Data

The record of monsoon onset dates at the Kerala coast in southwestern India as obtained from the India Meteorological Department is listed in Table 1 from 1958 to 1978. For the convenience of numerical treatment the onset date is expressed in terms of the number of days from 1 April in this study. The India Meteorological Department's 1967 Forecasting Manual lists the criteria by Ananthakrishnan *et al.* (see Subbaramayya and Kumar, 1978), according to which "beginning from 10 May if at least five out of seven stations report 24-hourly rainfall 1 mm or more for two consecutive days the forecaster should declare on the second day that the monsoon has advanced over Kerala." Despite the controversies among field observers for the precise determination of the monsoon onset date, such a definition of the monsoon onset date by rainfall is generally recognized as most definitive in southwestern India, and the uncertainty associated with the recorded onset dates is considered to be at most one or two days by the field meteorologists.

TABLE 1. Onset of monsoon at the Kerala coast.

Year	Month	Day	Number of days from 1 April
1958	June	14	75
1959	May	31	61
1960	May	14	44
1961	May	18	48
1962	May	17	47
1963	May	31	61
1964	June	6	67
1965	May	26	56
1966	June	1	62
1967	June	6	67
1968	June	8	69
1969	May	25	55
1970	May	26	56
1971	May	29	59
1972	June	18	79
1973	May	23	53
1974	May	26	56
1975	May	31	61
1976	June	1	62
1977	May	29	59
1978	May	28	58

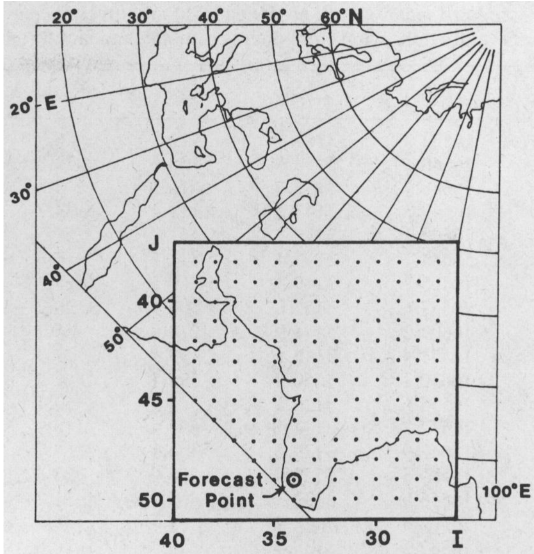

FIG. 1. NMC octagonal grids over Indian region and the forecast point.

The daily 1200 GMT geopotential height and temperature during April at 100 and 700 mb were obtained for a period from 1958 to 1978 from the National Meteorological Center's (NMC) octagonal grid data archives through the National Center for Atmospheric Research. As shown in Fig. 1, a grid point (NMC grid indices I = 34, J = 49) at 12.4°N and 76.5°E was designated as the forecast point, which is in the same general area as the Kerala coast. The eastward and northward components of winds at these levels were approximated by the geostrophic wind components at the forecast point from the NMC data set, and the kinetic energy was obtained with the approximated wind components. The daily values of these upper air parameters were averaged to obtain the respective April mean values. Because of a missing portion of the data the April mean upper air parameters in 1960 and 1961 could not be utilized, these two years are excluded from the total data

This content downloaded from 130.190.247.198 on Thu, 23 Feb 2023 11:10:21 UTC
All use subject to https://about.jstor.org/terms

TABLE 2. April mean values of 100 and 700 mb upper air parameters at the forecast point. The geopotential height is in units of m, temperature in K, wind components in m s^{-1}, and kinetic energy per unit mass in m^2 s^{-2}. The parameters are defined in Section 2.

Year	Z100	Z700	K100	K700	U100	U700	V100	V700	T100	T700
1958	16624.9	3162.6	112.6	46.9	7.3	−7.5	4.9	0.7	196.3	285.5
1959	16668.3	3158.6	135.6	74.0	−4.0	−7.6	−1.0	−3.7	195.0	285.7
1962	16638.7	3155.9	70.5	3.6	−1.0	0.9	2.9	0.4	201.1	281.8
1963	16549.3	3143.2	111.4	34.7	4.7	−1.2	2.9	−3.9	197.0	281.7
1964	16609.1	3150.8	511.7	50.5	−5.8	−3.7	−12.7	−4.1	195.3	285.8
1965	16684.1	3146.4	841.7	56.6	18.9	−5.6	−4.5	−1.0	197.7	284.9
1966	16705.2	3149.3	860.3	33.3	14.0	−4.4	−16.6	−2.7	197.6	285.6
1967	16557.6	3152.6	303.4	28.2	10.3	−2.2	−1.4	−4.8	195.7	287.3
1968	16609.2	3141.9	1270.8	55.7	30.3	−1.1	8.8	−0.9	195.9	284.7
1969	16531.4	3140.5	199.4	41.3	6.3	−3.8	−6.1	−4.8	195.9	286.9
1970	16630.1	3150.0	599.1	42.3	6.1	0.0	−3.3	0.7	195.5	284.2
1971	16555.6	3126.5	323.9	52.1	−2.2	−4.8	−8.8	−1.3	193.2	282.6
1972	16537.7	3134.5	503.4	110.9	6.1	−6.6	−8.4	−3.2	194.7	282.7
1973	16642.9	3139.2	386.1	67.7	3.6	−7.4	5.5	−3.3	195.2	284.9
1974	16585.6	3129.0	506.0	33.6	−3.0	−2.8	0.1	−0.5	193.5	283.2
1975	16556.9	3132.3	65.4	31.2	0.4	−6.3	0.5	−2.7	192.3	284.8
1976	16567.4	3142.4	91.7	35.7	−7.0	−7.0	−1.7	−1.7	196.7	283.3
1977	16582.5	3138.4	88.3	20.8	−2.7	−2.8	−3.6	−0.8	194.3	284.2
1978	16577.5	3143.3	62.9	29.5	−1.1	−5.5	−3.0	−2.1	198.2	284.8

period. Throughout this paper the following notation is adopted for the April mean values of 10 upper air parameters at the forecast point:

Z100 The geopotential height at 100 mb
Z700 The geopotential height at 700 mb
K100 The kinetic energy at 100 mb
K700 The kinetic energy at 700 mb
U100 The eastward component of wind at 100 mb
U700 The eastward component of wind at 700 mb
V100 The northward component of wind at 100 mb
V700 The northward component of wind at 700 mb
T100 The temperature at 100 mb
T700 The temperature at 700 mb.

3. April mean upper air parameters and scheme of regression analysis

The computed April mean values of the upper air parameters at 100 and 700 mb at the forecast point as listed in Table 2 indicate an obvious interannual variation of these parameters. In the case of winds, variation not only in magnitude but in sign is observed. The significant interannual variation of the upper air circulation prior to the monsoon onset is also confirmed by other investigators. For instance, Awade *et al.* (1978) computed the meridional transport of momentum by standing eddies for the strong monsoon year of 1967 and the weak monsoon year of 1972, and showed that the April values of transport at 20°N contrasted greatly. Winston and Krueger (1977) presented the relationship between the premonsoon radiative heating in spring over the Eurasian land complex and the monsoon onset, inferring the interannual variation of the upper air pattern prior to the monsoon onset.

The interannual variation of the April mean upper air parameters as indicated in Table 2 also can be observed in the daily fluctuation of these parameters as exemplified by the eastward component of the wind at 100 mb from 1 April to 30 June for the period from 1958 to 1968 as shown in Fig. 2. Although not presented in this paper, similar charts for other upper air parameters and for different years also indicate the significant interannual variation of the pattern prior to the monsoon onset.

Despite the obvious interannual variations of the antecedent upper air conditions, however, the pattern of variations of any single parameter in association with the monsoon onset date is by no means regular or clear. As shown in Table 3, the simple correlation between the recorded onset date and the single antecedent upper air parameter is significant only with K700 at the 5% level. If single regressions of the monsoon onset dates in Table 1 on single April mean upper air parameters in Table 2 are calculated, the corresponding coefficient of determination r^2 for these parameters shows low values (Table 4), indicating the limitation in the use of single antecedent upper-air parameters to relate with the monsoon onset date. This leaves the multi-regression as the preferred alternative in fitting the data in Tables 1 and 2.

The least-squares fitting of the monsoon onset date and the April mean values of upper air parameters are made to the general linear regression equation

$$y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_mx_m, \qquad (1)$$

where y is the monsoon onset date, b the regression coefficient, x the upper air parameter, and m the number of upper air parameters employed in the functional expression of y. For each of all available

This content downloaded from
130.190.247.198 on Thu, 23 Feb 2023 11:10:21 UTC
All use subject to https://about.jstor.org/terms

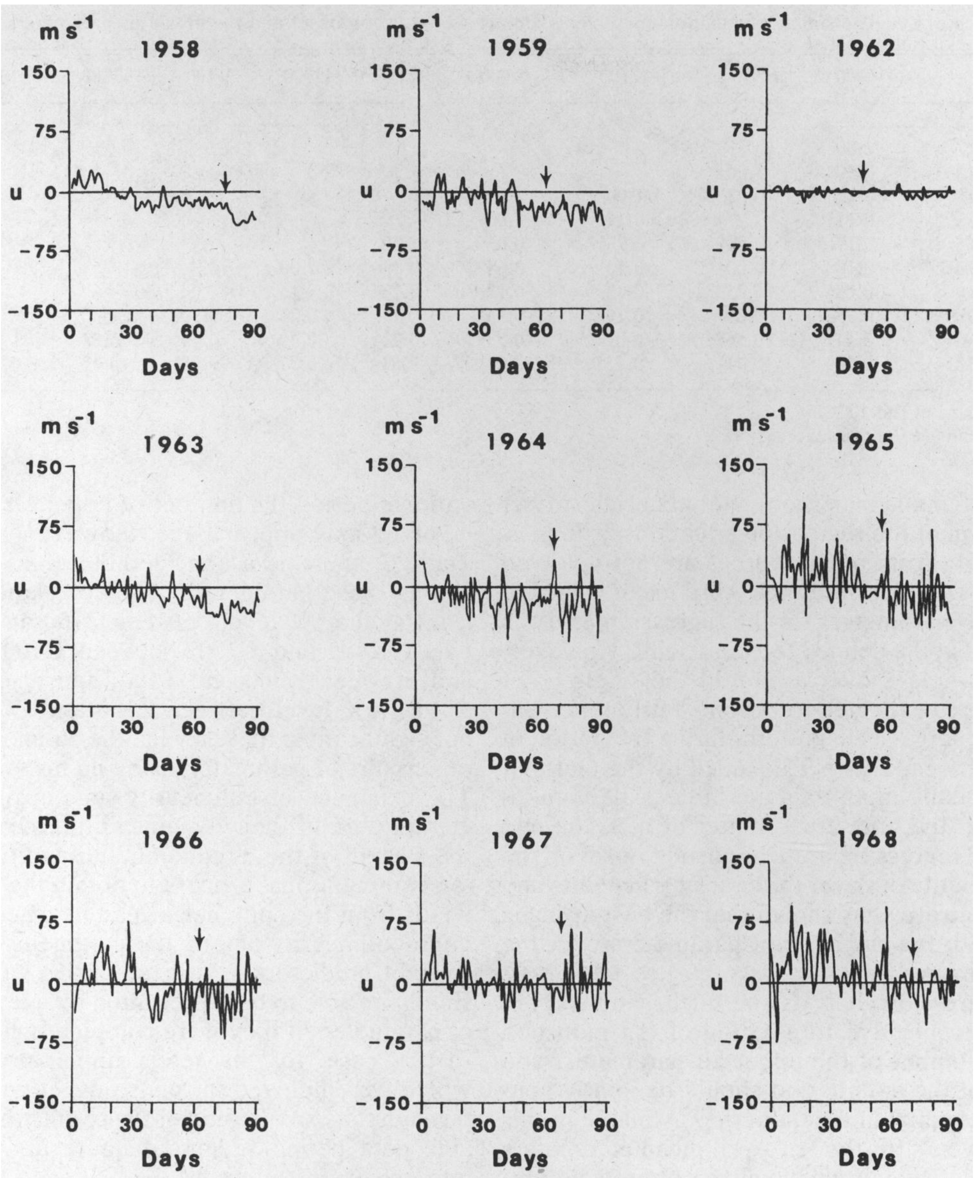

FIG. 2. The daily 100 mb eastward wind component at the forecast point for a 90-day period from 1 April, 1958–68.

combinations of up to ten upper air parameters, the fitting to Eq. (1) may be individually computed for various data periods up to 19 years from 1958 to 1978, excluding 1960 and 1961 when data are not usable. With the differences between the observed onset dates and fitted onset dates, the coefficient of multiple determination R^2 and the mean square error (MSE) as described in the standard texts on linear statistical models (e.g., Neter and Wasserman, 1974) are utilized to examine the regression fittings obtained. The program package of the Statistical Analysis System (Barr *et al.*, 1976) was employed in the computation of regression coefficients and statistical procedures.

Some general comments may be appropriate here concerning the multi-regression fitting of the monsoon onset date on the April mean upper air parameters and its subsequent application in an objective forecasting technique. The use of the NMC grid data over the forecast point involves the problem of a systematic bias inevitable for the grid interpolation due to the relatively sparse aerological observations in the tropical area of the Eastern Hemisphere. Further bias may also exist due to the computations of geostrophic winds in lower latitudes. However, the employment of the NMC grid data is necessary as the only available source of large-scale observational data for a reasonable length of time and these

This content downloaded from 130.190.247.198 on Thu, 23 Feb 2023 11:10:21 UTC
All use subject to https://about.jstor.org/terms

TABLE 3. The simple correlation coefficients among the monsoon onset and the upper air parameters.

	Onset	Z100	Z700	K100	K700	U100	U700	V100	V700	T100	T700
Onset	1.00										
Z100	−0.27	1.00									
Z700	0.10	0.50*	1.00								
K100	0.18	0.32	−0.23	1.00							
K700	0.53**	−0.11	−0.27	0.19	1.00						
U100	0.25	0.19	0.04	0.73**	0.06	1.00					
U700	−0.36	−0.18	0.00	0.13	−0.68**	0.13	1.00				
V100	−0.10	−0.05	0.15	0.12	0.12	0.23	0.16	1.00			
V700	−0.16	0.34	0.09	0.19	−0.25	0.14	0.24	0.32	1.00		
T100	−0.33	0.41	0.56**	−0.07	−0.53**	0.23	0.40	0.14	0.22	1.00	
T700	0.18	0.18	0.41	0.13	0.03	0.23	−0.31	−0.18	−0.42	−0.13	1.00

* Significant at the 10% level.
** Significant at the 5% level.

types of systematic bias should not materially affect the utilization of the regression equations as long as upper air data from other sources are not involved as predictors. In this study the April mean values of the upper air parameters for the regression analysis are obtained at the defined forecast point. However, by taking averages over a month, the single point parameters are expected to become fairly representative of the large-scale pattern during the period in the forecast area. This is also aided by the fact that the original daily input data are large-scale data interpolated at the grid. As a matter of fact, the test fitting of the regression of the monsoon onset on the upper air parameters over the nearby grid points and over the large area has shown that the best fitting is obtained with the single point parameters over the forecast point.

In the regression analysis and further in its application in an objective forecasting of the monsoon onset, the number of the upper air parameters and the length of the data period should be jointly considered. The maximum data period available in this study is 19 years for the ten April mean parameters in Table 2, of which eight will be chosen as predictors as will be discussed later in this paper. For the multi-regression analysis with this limited observational period the number of upper air parameters involved may appear large. However, as shown in Table 3, there is a considerable multi-collinearity among these parameters. The correlation is significant at the 5% level between Z700 and T100, between K100 and U100, between K700 and U700, and between K700 and T100, and also significant at the 10% level between Z100 and Z700. Though not significant at these levels, the collinearity among other pairs of parameters is by no means negligible. The existence of collinearity among these parameters is due to their dynamical interdependence in the system of the circulation and additionally from the computational process to obtain the geostrophic winds from the daily height pattern. Because of the multi-collinearity among the parameters, the effect of eight predictors as utilized in this study actually should be able to be represented by a lesser number of parameters if they were completely independent. In any case, for the set of upper air parameters which we employed in this study, it will be shown through the examination of the results that the available data period seems adequate for the specific purpose of this attempt.

TABLE 4. Coefficient of determination r^2 with single parameter regression equations based on data sets of 1958–74, 1958–76 and 1958–78.

Variables	1958–74	1958–76	1958–78
K700	0.3062	0.2921	0.2987
U700	0.1314	0.1165	0.1092
T100	0.1120	0.0832	0.0817
Z100	0.0709	0.0642	0.0531
U100	0.0630	0.0538	0.0637
T700	0.0342	0.0318	0.0302
K100	0.0324	0.0284	0.0392
V700	0.0250	0.0240	0.0271
V100	0.0099	0.0101	0.0088
Z700	0.0092	0.0090	0.0113

4. Regression fitting and forecasting of monsoon onset

When a set of regression coefficients is obtained for a particular combination of upper air parameters by least-squares fitting, this regression equation may yield the fitted onset dates for the data period with the April mean values of the upper air parameters of the respective years. Since up to ten upper air parameters are considered, various combinations of upper air parameters as regressors may be grouped into ten classes according to the number of regressors m. In each class of m from 1 to 10, all regression equations are compared within the class, and one equation is selected as the best combination of the upper air parameters to give the maximum R^2 and minimum MSE. The regressions thus selected for different numbers of regressors are listed in Table

This content downloaded from
130.190.247.198 on Thu, 23 Feb 2023 11:10:21 UTC
All use subject to https://about.jstor.org/terms

TABLE 5. Examination of the fitted date of monsoon onset with different number of variables for the period 1958–74.

Number of regressors (m)	R^2	MSE (days^2)	F ratio	P value	Best combination of upper air parameters by the maximum R^2 and minimum MSE values
1	0.306	54.85	5.74	0.0324*	K700
2	0.372	53.83	3.55	0.0616	Z700, K700
3	0.518	45.01	3.95	0.0390*	Z100, Z700, K700
4	0.737	27.04	7.00	0.0059**	Z100, Z700, K100, U700
5	0.855	16.58	10.60	0.0015**	Z100, Z700, K100, U700, T700
6	0.885	14.78	10.25	0.0022**	Z100, Z700, K100, U100, U700, T700
7	0.915	12.87	10.40	0.0031**	Z100, Z700, K100, U700, V100, T100, T700
8	0.943	9.69	12.50	0.0032**	Z100, Z700, K100, K700, U700, V100, T100, T700
9	0.947	10.54	10.28	0.0097**	Z100, Z700, K100, K700, U700, V100, V700, T100, T700
10	0.949	13.15	7.42	0.0342*	Z100, Z700, K100, K700, U100, U700, V100, V700, T100, T700

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

5 for the period from 1958 to 1974 in terms of the selected combination of upper air parameters, R^2 and MSE with its F-ratio and P-value.

In observing Table 5 it appears that the fitting of the onset date significantly improves when m increases to 5. P-values in Table 5 show that the F-ratio associated with MSE is significant at the 1% level in the range of m from 4 to 9. R^2 increases as the number of regressors increases, and after m reaches 8 the increase in R^2 becomes essentially asymptotic. Therefore, the eight upper air parameters specified in Table 5 for the case of $m = 8$ are taken as the best combination of regressors to represent April upper air conditions preceding the monsoon onset. As seen in Table 5, the addition of V700 in the class of $m = 9$ and that of U100 and V700 for $m = 10$ does not improve the fitting. It is probable that the utility of U100 and V700 is essentially represented by K100 and K700 in terms of the strength of the flow. It is noted that for $m \geqslant 5$ the geopotential height at 100 and 700 mb, the strength of the flow at 100 and 700 mb in terms of either kinetic energy or wind components, and the low-level temperature in terms of T700 are always involved as regressors. It then appears that the multi-regression fitting of the monsoon onset seems at its optimum when the regression coefficients are computed for the regression equation

$$Y = b_0 + b_1\mathrm{Z100} + b_2\mathrm{Z700} + b_3\mathrm{K100} + b_4\mathrm{K700} + b_5\mathrm{U700} + b_6\mathrm{V100} + b_7\mathrm{T100} + b_8\mathrm{T700}, \quad (2)$$

where Z100 and Z700 are in meters, K100 and K700 in m^2 s^{-2}, U700 and V100 in m s^{-1} and T100 and T700 in K. The regression coefficients $b_0, b_1, \ldots, b_8$ are to be obtained by the least-squares fitting of the onset date and the April mean upper air parameters for a specified period of time. In illustrating the selection of these eight parameters in Table 5, a 15-year data period from 1959 to 1974 was utilized. All other tested data periods from 12 years to 19 years gave very similar results.

The regression equation (2) was fitted separately for eight data periods from 12 years (1958–71) to 19 years (1958–78), and the fitted dates of monsoon onset are compared with the recorded dates in Table 6. The onset dates listed below the stepped line in Table 6 are for years whose upper air data are not involved in the least-squares fitting of the respective regression equations. Therefore, these values may be regarded as the independent forecasts. As is immediately seen, the fitted onset dates during the data periods compare very closely with the dates that were actually recorded. The average deviation of the fitted onset date from the recorded onset date is only 1.2 days for the 1958–71 period with a small increase to 2.3 days for the 1958–77 period. The forecasted dates beyond the data period show a fairly close prediction of the recorded onset dates, especially when more than 13 years of data period is involved in the least-squares fitting of the regression coefficients. With the data periods of 1958–71 through 1958–77, the average deviations of the forecasted onset date for years beyond the data periods are in the range 4–7 days from the recorded onset date, mostly centering around 4 or 5 days. It should be specifically noted here that the increase of the data sample through extension of the data period in approximating the regression equations does not seem to improve the regression fitting or forecasting of the onset materially for the upper air parameters employed as the predictors.

The utility of the multi-regression equation (2) in forecasting the monsoon onset was tested in another way as seen in Table 7, the onset dates being forecasted for the data periods 1959–74 (case A) and 1959–78 (case B). Each of the forecasted onset dates in Table 7 was obtained with separate regression equations whose coefficients were fitted without involving the data from the forecast year, thus

This content downloaded from
130.190.247.198 on Thu, 23 Feb 2023 11:10:21 UTC
All use subject to https://about.jstor.org/terms

TABLE 6. Fitting and forecasting of monsoon onset date by regression equations approximated with different data periods. Dates listed below the stepped line are independent forecasts with the regression equations. The differences between the recorded and fitted dates are shown in parentheses.

Year	Recorded onset date	Data period to approximate regression equations							
		1958–71	1958–72	1958–73	1958–74	1958–75	1958–76	1958–77	1958–78
1958	75	76 (−1)	76 (−1)	76 (−1)	76 (−1)	76 (−1)	73 (2)	73 (2)	73 (2)
1959	61	59 (2)	60 (1)	61 (0)	61 (0)	61 (0)	63 (−2)	63 (−2)	63 (−2)
1962	47	47 (0)	47 (0)	47 (0)	46 (1)	46 (1)	45 (2)	46 (1)	47 (0)
1963	61	59 (2)	61 (0)	61 (0)	61 (0)	61 (0)	60 (1)	61 (0)	61 (0)
1964	67	67 (0)	70 (−3)	68 (−1)	68 (−1)	68 (−1)	68 (−1)	69 (−2)	69 (−2)
1965	56	58 (−2)	59 (−3)	60 (−4)	60 (−4)	60 (−4)	59 (−3)	58 (−2)	59 (−3)
1966	62	60 (2)	59 (3)	59 (3)	59 (3)	59 (3)	58 (4)	58 (4)	58 (4)
1967	67	66 (1)	64 (3)	64 (3)	64 (3)	64 (3)	64 (3)	65 (2)	65 (2)
1968	69	68 (1)	68 (1)	69 (0)	69 (0)	69 (0)	69 (0)	69 (0)	68 (1)
1969	55	56 (−1)	56 (−1)	57 (−2)	57 (−2)	57 (−2)	57 (−2)	58 (−3)	59 (−4)
1970	56	58 (−2)	57 (−1)	57 (−1)	57 (−1)	57 (−1)	60 (−4)	60 (−4)	60 (−4)
1971	59	59 (0)	58 (1)	60 (−1)	61 (−2)	62 (−3)	61 (−2)	62 (−3)	62 (−3)
1972	79	65 (14)	77 (2)	77 (2)	77 (2)	76 (3)	75 (4)	75 (4)	75 (4)
1973	53	49 (4)	46 (7)	51 (2)	51 (2)	51 (2)	52 (1)	52 (1)	53 (0)
1974	56	56 (0)	50 (6)	52 (4)	54 (2)	55 (1)	55 (1)	56 (0)	56 (0)
1975	61	62 (−1)	54 (7)	57 (4)	59 (2)	60 (1)	59 (2)	60 (1)	60 (1)
1976	62	70 (−8)	70 (−8)	70 (−8)	71 (−9)	72 (−10)	67 (−5)	67 (−5)	67 (−5)
1977	59	55 (4)	48 (11)	50 (9)	51 (8)	52 (7)	52 (7)	53 (6)	54 (5)
1978	58	56 (2)	56 (2)	57 (1)	57 (1)	57 (1)	54 (4)	54 (4)	56 (2)

making the forecast independent of the fitting of regression coefficients. The forecasted onset dates thus obtained are also recognized as being fairly close to the recorded onset dates. The average deviations of the forecasted date from the recorded date are 4.6 days for both data periods of 1958–74 and 1958–78. This is about the same deviation as shown in the forecast beyond the data period in Table 6. The comparison of case A and case B as shown in Table 7 again indicates that the extended data period in fitting the regression coefficients for the case B group does not seem to have improved forecasting materially.

TABLE 7. Forecast experiment for the period 1958–74 (case A) and 1958–78 (case B) with regression equations which are obtained without upper air information of the forecast year. The differences between the recorded and forecasted dates are shown in parentheses.

Year	Recorded onset date from 1 April	Forecasted onset date from 1 April	
		Case A	Case B
1958	75	78 (−3)	70 (5)
1959	61	62 (−1)	65 (−4)
1962	47	44 (3)	46 (1)
1963	61	62 (−1)	61 (0)
1964	67	69 (−2)	69 (−2)
1965	56	62 (−6)	60 (−4)
1966	62	53 (9)	52 (10)
1967	67	62 (5)	64 (3)
1968	69	71 (−2)	66 (3)
1969	55	62 (−7)	63 (−8)
1970	56	59 (−3)	62 (−6)
1971	59	64 (−5)	63 (−4)
1972	79	65 (14)	62 (17)
1973	53	49 (4)	53 (0)
1974	56	52 (4)	56 (0)
1975	61		60 (1)
1976	62		71 (−9)
1977	59		52 (7)
1978	58		54 (4)

Fig. 3 compares recorded, fitted and forecasted monsoon onset dates for the period from 1958 to 1978. Fitted dates are by the least-squares fitting of Eq. (2) with the 1958–78 data period as shown in Table 6, and the forecasted dates are the results of an independent forecast with regression equations for the 1958–78 period which are obtained without upper air information of the forecast year (case B of Table 7). A generally good correspondence of the forecasted date with the recorded and fitted dates may be noted except for a marked difference between the forecasted and onset dates for 1972 and to a lesser degree for 1966 and 1976. Considering the results of forecasting the onset date as shown in Tables 6 and 7 and Fig. 3, it may be stated that the multi-regression scheme with the April mean values of upper air parameters predicts the monsoon onset reasonably well about one month in advance of the onset in regular years, about a half-month in advance in very early monsoon years, and about one and a half months in advance in very late monsoon years.

It should be noted in Table 7 and Fig. 3 that in 1972 when the monsoon failure is indicated by the ex-

This content downloaded from
130.190.247.198 on Thu, 23 Feb 2023 11:10:21 UTC
All use subject to https://about.jstor.org/terms

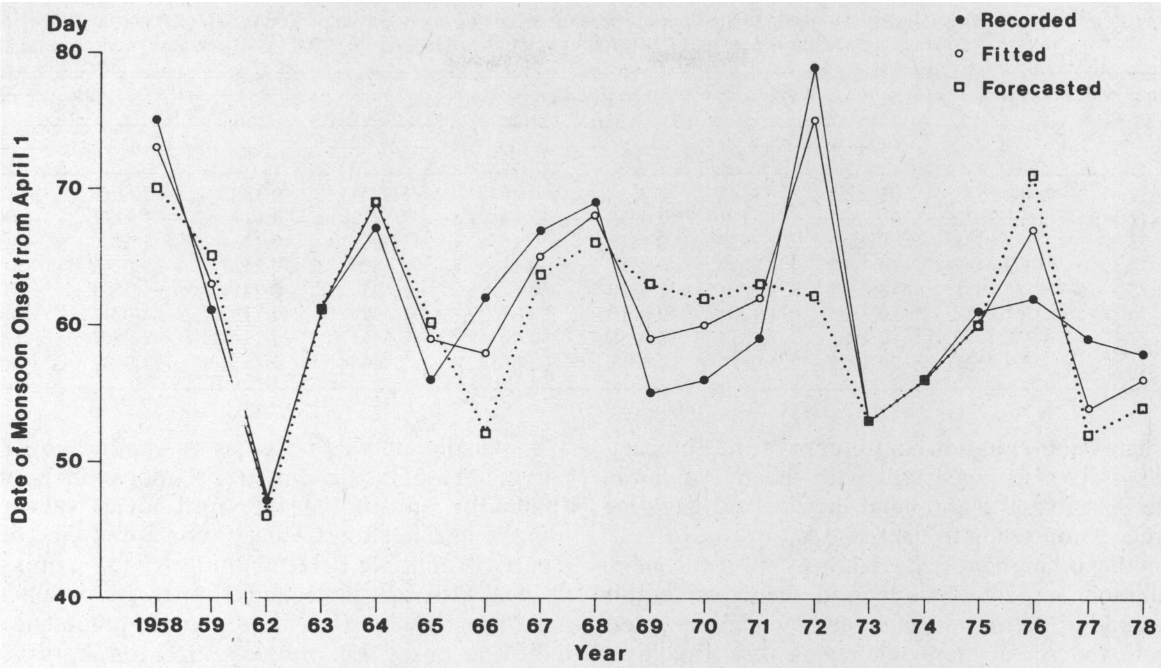

FIG. 3. The recorded, fitted and forecasted monsoon onset dates from 1958 to 1978.

tremely late arrival of the monsoon the forecasted date leads the recorded date by 17 days in the case B forecast (with 1958–78 data base) and by 14 days in the case A forecast (with 1958–74 data base). The discrepancy of the forecasted data in 1972 cannot be explained by the inability of the regression equation to forecast in the extreme range since the forecasts in the very late monsoon year of 1958 and very early monsoon year of 1962 do not show this discrepancy. The discrepancy in 1972 therefore seems to be found in the specific conditions associated with the 1972 monsoon failure and the inadequacy of the scheme to incorporate this information. The anomalous large-scale circulation patterns in association with the 1972 monsoon failure have been examined from various aspects (e.g., Kanamitsu and Krishnamurti, 1978; Krueger and Winston, 1975; Murakami, 1975), and are shown to be extraordinary in the extent the anomalies were observed over the global scale of the tropical belt. In Table 2, 1972 April upper air conditions are specifically marked by a large K700. Since U700 and V700 are not particularly large it should mean that strong and variable winds dominated in the lower troposphere. However, it is difficult to ascertain the definitive pre-monsoon upper air conditions associated with the late monsoon from the April mean conditions listed in Table 2; indeed this difficulty associated with the complex nonlinear physical system is the reason the multi-regression approach has been adopted in this study. In any case, it may be stated that the independent forecast for 1972 failed because the regression equation utilized has not incorporated the specific condition which existed in 1972.

5. Discussion

In this exploratory attempt at monsoon onset forecasting, no definitive multi-regression equation is uniformly applied to yield all independent forecasts. Instead, separate regression equations are approximated for different data periods without involving the data of the forecast year. Because of the limited time period of less than 20 years for usable upper air data, the derivation of a stabilized regression equation is not warranted. For example, Table 8 lists the eight regression equations approximated for eight different data periods, 1958–71 to 1958–78, which are utilized in the onset data fitting and forecasting as shown in Table 6. It is readily seen that the regression coefficients of some predictors show considerable variations by extending the limit of the data period from 1971 to 1975. For the 1958–76, 1958–77 and 1958–78 periods the variation is stabilized considerably, but it will be reasonable to assume that a much longer data period is needed to achieve the stability of the regression equation. Nevertheless, it should be noted that even without a stabilized regression equation the individually approximated regression equations for different data periods are shown to give compatible predictions (see Tables 6 and 7). As noted previously, there is also no specific indication in observing Tables 6 and 7 that the extended data period in approximating

This content downloaded from
130.190.247.198 on Thu, 23 Feb 2023 11:10:21 UTC
All use subject to https://about.jstor.org/terms

TABLE 8. Regression equations approximated with different periods of upper air data: $Y = b_0 + b_1Z100 + b_2Z700 + b_3K100 + b_4K700 + b_5U700 + b_6V100 + b_7T100 + b_8T700$.

Data period	Regression coefficients								
	b_0	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8
1958–71	754.8512	−0.2022	1.3246	0.0327	−0.2870	−3.4954	−0.1882	−2.6010	−3.5416
1958–72	1021.3642	−0.2564	1.5158	0.0356	−0.1321	−3.5130	−0.4188	−1.4372	−4.2592
1958–73	982.0733	−0.2248	1.2978	0.0316	−0.1056	−3.3405	−0.2733	−1.3270	−3.6271
1958–74	1094.4483	−0.2207	1.2692	0.0317	−0.1296	−3.3829	−0.2527	−1.6243	−3.7357
1958–75	1156.4911	−0.2221	1.2684	0.0319	−0.1458	−3.4934	−0.2454	−1.7389	−3.7822
1958–76	761.8937	−0.1885	1.1704	0.0276	−0.0707	−2.5395	−0.2193	−2.0278	−3.0655
1958–77	736.1979	−0.1783	1.1314	0.0256	−0.0774	−2.3455	−0.2342	−2.2582	−2.9769
1958–78	791.4521	−0.1798	1.0876	0.0251	−0.0782	−2.4068	−0.2310	−1.9641	−2.8005

the regression equation may improve the forecasting. This is also consistent with the R^2 values in Table 9 in that the extended data period in fitting the regression tends to decrease R^2.

On the other hand, the patterns of the general circulation are expected to undergo considerable variations in the period of a few decades, in reference to the Asian monsoon region (see Tsuchiya, 1971), and the constantly revised regression with the data of an adequate length of period will be preferable to a regression equation from the data sample of a very long period. Therefore, it seems feasible to make the continuous updating of the regression equation with the annual addition of the new data an integral part of the forecasting scheme, although the actual method of revising the regression coefficients and the proper length of the data period are yet to be determined.

The regression equation is approximated in this study with April mean upper air parameters obtained from the daily NMC octagonal grid data at the selected forecast point. Whereas a similar functional relationship is expected to be derived at other forecast points with a similar set of data, the regression should be fitted at individual points for individual sources of data.

TABLE 9. Coefficient of multiple-determination R^2 for regression fittings of Eq. (2) with upper air parameters averaged over various pre-monsoon periods.

Pre-monsoon period	Data period to approximate regression equations				
	1958–74	1958–75	1958–76	1958–77	1958–78
1 Apr–30 Apr	0.94	0.94	0.90	0.86	0.86
1 Apr–15 Apr	0.46	0.45	0.45	0.42	0.42
15 Mar–15 Apr	0.35	0.37	0.37	0.36	0.36
15 Mar–30 Mar	0.38	0.37	0.37	0.33	0.32
1 Mar–30 Mar	0.55	0.55	0.55	0.51	0.52
1 Mar–15 Mar	0.35	0.33	0.33	0.28	0.28
15 Feb–15 Mar	0.33	0.32	0.32	0.26	0.24
1 Feb–28 Feb	0.33	0.33	0.32	0.31	0.31

A question may be raised as to whether upper air parameters of some earlier pre-monsoon periods should be substituted for April mean values as utilized in this study. Table 9 compares the coefficients of multiple determination R^2 for regression fittings with the same eight upper air parameters averaged over various time periods preceding the monsoon onset. It is obvious that the April mean values as used in this study give the closest fitting of the monsoon onset dates as indicated by high R^2 values. On the contrary the R^2 value drops sharply with mean upper air parameters of other pre-monsoon periods. It may be seen that we do not gain by using the upper air parameters of earlier periods beyond the April mean values.

In this paper the 100 and 700 mb parameters at the forecast point are employed in formulating the forecasting scheme. However, it is easily seen that other meteorological parameters which are related to the development of the monsoon circulation system could be added or partially substituted as viable predictors. Parameters to describe the overall large-scale circulation pattern in the area, to relate the Asian summer monsoon to the event of the distance, or to measure the large-scale heating pattern may come under consideration in this regard. For instance, Ananthakrishnan (1970, 1977) has shown that the seasonal reversal of the pressure gradients through the lower and upper levels is a distinct feature in months preceding the monsoon onset. Since the time of Walker (1923, 1924), who measured the Southern Oscillation in terms of the pressure difference between the southeastern Pacific and Indonesia, there have been many attempts to relate the Asian summer monsoon to the fields of global-scale circulation. Recently, Ramaswamy and Pareek (1978) illustrated the potential forecasting value of the association between the Asian summer monsoon activities and the upper air flow patterns in the Southern Hemisphere between 50 and 160°E. Concerning the heating variables it is particularly noteworthy that the analyses by Hahn and Shukla (1976) and Winston and Krueger (1977) and the

This content downloaded from
130.190.247.198 on Thu, 23 Feb 2023 11:10:21 UTC
All use subject to https://about.jstor.org/terms

numerical experiment by Hahn and Manabe (1976) indicate the possible use of snow cover over Eurasia as a parameter in forecasting the monsoon. Winston and Krueger attribute the relatively weak 1976 monsoon to the high April albedos over the Eurasian mountain complex including Soviet central Asia and the northwest Himalayas. They further point out that the Eurasian snow cover preceding the 1972 monsoon was the greatest of the decade. If these parameters are properly quantified in the multi-regression equation to describe the patterns of the heating and the large-scale circulation, it is conceivable to extend the forecast range of the monsoon onset as well as to improve the forecast by reducing the possibility of the large deviation such as in the case of the 1972 onset date.

6. Concluding remarks

It has been shown that, through the use of the April mean values of daily upper air parameters at the selected point, the recorded onset dates of the summer monsoon in southwestern India can be closely related functionally to the antecedent upper air conditions. It has also been demonstrated by various uses of the data sample in approximating the regression equations that such a scheme of regression analysis can be applied to the objective forecast of the monsoon onset with a reasonable accuracy. Because of the time variations of the patterns of the general circulation, the constant updating of the regression equation should be an integral part of the forecasting procedure if this type of scheme is utilized.

Other meteorological parameters pertaining to the large-scale heating and development of the global-scale circulation should be considered in the future as viable additions or partial substitutions to the predictors employed in this study. The possible result of this would be the improvement of accuracy and the extension of the forecast range.

Acknowledgments. The authors are indebted to Dr. Sharon LeDuc and Dr. L. T. Steyaert of the Center for the Environmental Assessment Services/NOAA for the consultation of the statistical techniques. The consultation of Drs. P. K. Das, R. P. Sarker and V. Thapliyal of the India Meteorological Department and Dr. D. R. Sikka of the Indian Institute of Tropical Meteorology on various aspects of the monsoon onset is gratefully acknowledged. The authors also appreciate the technical assistance rendered by C. W. Irwin, J. M. Makilya and Dawn Stringfield, and the reading of the manuscript by H. A. Burgdorf.

REFERENCES

Ananthakrishnan, R., 1970: Reversal of pressure gradients and wind circulation across India and the southwest monsoon. *Quart. J. Roy. Meteor. Soc.*, **96**, 539–542.

——, 1977: Some aspects of the monsoon circulation and monsoon rainfall. *Pure Appl. Geophys.*, **115**, 1209–1249.

Awade, S. T., K. S. R. Rao, and G. C. Asnani, 1978: The meridional transport of momentum in the wave number regime during the contrasting monsoon years. *Indian J. Meteor. Hydrol. Geophys.*, **29**, 283–292.

Barr, A. J., J. H. Goodnight, G. P. Sall, and J. T. Helwig, 1976: *A User's Guide to SAS-76*. SAS Institute, Inc., Raleigh, NC, 329 pp.

Findlater, J., 1969: A major low-level air current near the Indian Ocean during the northern summer. *Quart. J. Roy. Meteor. Soc.*, **95**, 362–380.

Flohn, H., 1968: Contribution to a meteorology of the Tibetan Highlands. Atmos. Sci. Pap., Colorado State University, [NTIS PB-182-225].

Hahn, D. G., and S. Manabe, 1975: The role of mountains in the south Asian monsoon circulation. *J. Atmos. Sci.*, **32**, 1515–1541.

——, and J. Shukla, 1976: An apparent relationship between Eurasian snow cover and Indian monsoon rainfall. *J. Atmos. Sci.*, **33**, 2461–2462.

Kanamitsu, M., and T. N. Krishnamurti, 1978: Northern summer tropical circulations during drought and normal rainfall months. *Mon. Wea. Rev.*, **106**, 331–347.

Krishnamurti, T. N., 1971: Observational study of tropical upper atmospheric motion field during northern hemisphere summer. *J. Appl. Meteor.*, **10**, 1066–1096.

Krueger, A. F., and J. S. Winston, 1975: Large-scale anomalies over the tropics during 1971–72. *Mon. Wea. Rev.*, **103**, 465–473.

Lockwood, J. G., 1974: *World Climatology*. St. Martin's Press, 330 pp.

Lorenz, E. N., 1956: Empirical orthogonal functions and statistical weather prediction. Dept. Meteor., MIT, Sci. Rep. No. 1, Contract AF19(604)1566, 49 pp. [Library of Congress, PB-126-724.].

Manabe, S., D. G. Hahn and J. L. Holloway, 1974: The seasonal variation of the tropical circulation as simulated by a global model of the atmosphere. *J. Atmos. Sci.*, **31**, 43–83.

Murakami, T., 1975: Interannual cloudiness changes. *Mon. Wea. Rev.*, **103**, 996–1006.

——, 1976: Cloudiness fluctuations during the summer monsoon. *J. Meteor. Soc. Japan*, **54**, 175–181.

Namias, J., 1959: Recent seasonal interactions between North Pacific waters and the overlying atmospheric circulation. *J. Geophys. Res.*, **64**, 631–646.

——, 1973: Response of the equatorial countercurrent to the subtropical atmosphere. *Science*, **181**, 1245–1247.

——, 1976: Some statistical and synoptic characteristics associated with El Niño. *J. Phys. Oceanogr.*, **6**, 130–138.

——, 1978: Multiple causes of North American abnormal winter 1976–77. *Mon. Wea. Rev.*, **106**, 279–295.

Neter, J., and W. Wasserman, 1974: *Applied Linear Statistical Models–Regression, Analysis of Variance and Experimental Designs*. Richard D. Irwin, Inc., Homewood, IL, 842 pp.

Ramage, C. S., 1971: *Monsoon Meteorology*. Academic Press, 296 pp.

——, and C. R. V. Raman, 1970: *Meteorological Atlas of the International Indian Ocean Expedition*, Vol. 2, *Upper-air*. National Science Foundation, Washington, DC., 121 charts. [Superintendent of Documents, Stock No. 3800-00124].

Ramaswamy, C., and R. S. Pareek, 1978: The southwest monsoon over India and its teleconnections with the middle and upper tropospheric flow patterns over the Southern Hemisphere. *Tellus*, **30**, 126–135.

Shukla, J., 1975: Effect of Arabian sea-surface temperature anomaly on Indian summer monsoon—A numerical experiment with the GFDL model. *J. Atmos. Sci.*, **32**, 503–511.

Sikka, D. R., 1977: Some aspects of the life history, structure

This content downloaded from
130.190.247.198 on Thu, 23 Feb 2023 11:10:21 UTC
All use subject to https://about.jstor.org/terms

and movement of monsoon depressions. *Pure Appl. Geophys.*, **115**, 1501–1529.

Singh, S. V., D. A. Mooley, and R. H. Kripalani, 1978: Synoptic climatology of the daily 700 mb summer monsoon flow patterns over India. *Mon. Wea. Rev.*, **106**, 510–525.

Staff Members of Academia Sinica, Peking, 1958: On the general circulation over Eastern Asia (II). *Tellus*, **10**, 58–75.

Subbaramayya, I., and O. S. R. U. Kumar, 1978: The onset and the northern limit of the southwest monsoon over India. *Meteor. Mag.*, **107**, 37–48.

Tsuchiya, I., 1971: Fluctuations of rainfall in southeast Asia—Equatorial Pacific and low and middle latitude circulations in the Southern Hemisphere. *Water Balance of Monsoon Asia*, M. M. Yoshino, Ed., University of Tokyo Press and University of Hawaii Press, 217–240.

Walker, G. T., 1923: Correlation in seasonal variations of weather, VIII: A preliminary study of world weather. *Mem. Indian Meteor. Dept.*, **24**, 75–131.

——, 1924: Correlation in seasonal variations of weather, X: Applications to seasonal forecasting in India. *Mem. Indian Meteor. Dept.*, **24**, 333–345.

Winston, J. S., and A. F. Krueger, 1977: Diagnosis of the satellite-observed radiative-heating in relation to the summer monsoon. *Pure Appl. Geophys.*, **115**, 1131–1144.

Yeh, T. -C., and C. -C. Chang, 1974: A preliminary experimental simulation on the heating effect of the Tibetan Plateau on the general circulation over eastern Asia in summer. *Sci. Sinica*, **17**, 397–420.

Yoshino, M. M., Ed., 1971: *Water Balance of Monsoon Asia*. University of Tokyo Press and University of Hawaii Press, 308 pp.

This content downloaded from
130.190.247.198 on Thu, 23 Feb 2023 11:10:21 UTC
All use subject to https://about.jstor.org/terms