Potential Predictability of Lower-Tropospheric Monsoon Circulation and Rainfall over India

S. V. SINGH AND R. H. KRIPALANI

Indian Institute of Tropical Meteorology, Pune-411 008, India

(Manuscript received 26 January 1985, in final form 30 October 1985)

ABSTRACT

The potential predictability of the lower tropospheric circulation and the rainfall over India during the peak summer monsoon season (July–August) is studied by analyzing the signal-to-noise ratio. Daily 700-mb heights, mean sea level pressure anomaly and rainfall at 220 stations for 21, 30 and 19 years, respectively, are used to represent the circulation and rainfall fields. The predictability of the circulation fields in general increases with decreasing latitude but is low over the area normally occupied by the monsoon trough. The potential predictability of rainfall is about 50% over the major parts of the country.

1. Introduction

The potential predictability of monthly or seasonal averages has been studied by comparing the interannual variability of the observed monthly or seasonal means to the natural variability of the time averages of day-to-day values. Pioneering work in this field has been done by R. A. Madden (1976, 1981). A similar approach was used by Shukla and Gutzler (1983) to study the predictability of seasonal northern hemispheric 500-mb contour heights and by Nicholls (1981, 1983) to study the temperature and pressure over the Australian region. However, Trenberth (1984a,b) recently provided a comprehensive review and critique of the various methods used in the past by the above (and other) authors to assess the persistence in day-to-day weather and the potential predictability of time-averaged means. He then proposed and employed an improved method for studying the potential predictability of geopotential heights over the Southern Hemisphere (Trenberth, 1985), the basic principles remaining the same. One of the principal results of these studies is that the predictability is dependent on season and geographical location, being higher for the tropics than for the middle or higher latitudes.

The southern limit for most of these studies has been 20°N. Availability of sufficient length of daily 700-mb contour height and mean sea level pressure anomaly grid point data over the Indian region prompted us to analyze the potential predictability of seasonal mean daily fields over the Indian region. We believe that this is the best available data set of the above fields over the Indian region and extending to 5°N, encompassing all synoptic activity taking place in this region. We followed the approach of Shukla and Gutzler (1983) or method A of Trenberth (1984a), which is based on simple analysis of variance for studying the potential predictability of bimonthly (July–August) means of 700-mb contour heights and mean sea level pressure anomaly. The potential predictability of July–August means of rainfall amounts over the homogeneous meteorological regions has been studied by following an approach similar to that of Katz (1983). The choice of averaging over the two months, July and August, was determined by the following factors: (i) this period is sufficiently long to be beyond the limit of day-to-day predictability; (ii) the meteorological fields and the weather remain almost stationary during this period of established monsoon; and (iii) external conditions such as sea surface temperature remain more or less constant.

2. Data

The daily (average of 0000 and 1200 GMT values) 700-mb contour heights are available for a period of 21 years (1958–78). A single chart consists of 98 points on a 2.5° lat/long diamond grid and covers the area between 5° to 35°N latitude and 60° to 100°E longitude. The contour height values were picked from the manually analyzed charts for 0000 and 1200 GMT and averaged. The daily average values were replotted and reanalyzed to determine whether the contour height fields faithfully reproduced the major circulation features observed on the original weather charts. The daily values of mean sea level pressure anomaly were available for a 30-year period (1946–75). Seventy-two 2.5° lat/2.5° long blocks covered the entire land area of the country and the Bay Islands. The daily rainfall data of 220 stations uniformly distributed over the country were obtained from the India Meteorological Department for a period of 19 years (1955–73). Simple
averages of these station data were taken to determine the daily spatially-averaged values for the 33 meteorological regions (see Fig. 3) covering the entire land area of the country including the two groups of islands: one in the Bay of Bengal and another in the Arabian Sea.

3. Methodology

We used principles similar to those of Madden (1976, 1981) to study the potential predictability of bimonthly means of the circulation fields. The method is identical to that of Shukla and Gutzler (1983) or method A of Trenberth (1984a). The interannual variability of a variable $x$ is estimated by

$$\sigma^2_{T,obs} = \frac{1}{n \nu_1} \left( \sum_{j=1}^{J} x_{ij}^2 - \frac{1}{n} \left( \sum_{j=1}^{J} x_{ij} \right)^2 \right),$$

(1)

and the natural variability of time average, i.e., the noise, is estimated by

$$\sigma^2_{T,noise} = \frac{1}{n \nu_2} \left( \sum_{j=1}^{J} \sum_{i=1}^{n} x_{ij}^2 - \sum_{j=1}^{J} \left( \sum_{i=1}^{n} x_{ij} \right)^2 \right),$$

(2)

where the subscript $i$ refers to the days in a season; the subscript $j$ to the seasons or years; the symbol $T$ refers to the length of time averaging or the number of days ($n$) in a season (equal to 62 in the present study); $\nu_1$ and $\nu_2$, the degrees of freedom used in the computation of the two variances, are given by

$$\nu_1 = \left\{ \begin{array}{ll} J - 1, & J \text{ being the number of years}, \\ 20 & \text{for the 700 mb contour height fields}, \\ 29 & \text{for the mean sea level pressure} \\ \text{anomaly fields}, \end{array} \right. \nu_2 = \frac{(n - 1)}{\tau_0} J,$$

(3)

in which $\tau_0$, often referred to as the effective time between independent observations (Madden, 1976), is estimated from the daily observations in the manner described below. The ratio between the two estimates of the interannual variances $\sigma^2_{T,obs}$ and $\sigma^2_{T,noise}$ is then tested for significance by the $F$-test:

$$F(\nu_1, \nu_2) = \frac{\sigma^2_{T,obs}}{\sigma^2_{T,noise}},$$

(4)

under the null hypothesis that they estimate the same uncorrelated interannual variance. Statistical tables can then be consulted for assessing the statistical significance. Since the $\sigma^2_{T,obs}$ includes the variance due to the noise and the sought signal, the factor ($F - 1$) can be considered as a measure of the signal-to-noise ratio.

The $\tau_0$ required in (3) is computed by

$$\tau_0 = 1 + \sum_{l=1}^{n} 2 \left( 1 - \frac{l}{n} \right) r(l),$$

(5)

(e.g., see Trenberth, 1985) where $r(l)$ are the autocorrelations for different lags. The upper limit of the summation was set to 15, after some experimentation. Thiebaux and Zwiers (1984) have noted that inclusion of higher lags in the summation increases the variability of $\tau_0$, as some values of $r(l)$ are substituted in place of zero or near-zero theoretical values. We observed that the autocorrelation decreases to insignificant value after lag 15, and hence the exclusion of the higher lag terms in (5) is justified. The autocorrelations needed in (5) are first determined for each year separately by using the time series of length 62 days. The $J$ autocorrelations for respective lags are then averaged after finding the Fisher’s $Z$-transformation which tends to normalize the distribution of correlations. It may be noted that Trenberth (1984a) has pointed out that this procedure, which uses individual year time-averages for computation, introduces some negative bias in the autocorrelations. This bias increases with increasing $r(l)$ and decreasing $T$. As we shall see below, the $r$ in the present study are $\approx 0.7$, and $T$ is 62. So, considering Figs. 1 and 2 of Trenberth (1984a) for $r(l) = 0.7$ and $T = 90$ and 30, respectively, we feel that the bias may not be a serious problem. As noted already for $r \approx 0.7$, the bias will be less than that shown in the above figures of Trenberth.

Since the rainfall is an intermittent variable, a different procedure is followed for determining the appropriate natural variability for the rainy days and the rain amounts (Katz, 1983). The method for determining the observed interannual variability remains the same.

The natural variability (noise) of rainfall averaged over $T$ days is given by (see Katz, 1983)

$$\sigma^2 = \frac{1}{T} p \sigma^2 + \frac{1}{T} (1 - p) \mu^2 (1 + d)/(1 - d),$$

(6)

where $p$ is the climatic or stationary probability of a rainy day, $\mu$ is the average amount of rain on a rainy day, $\sigma^2$ is the variance of the amount of rain received on rainy days only, and $d = \rho_{11} - \rho_{01}$ is a persistence factor, similar to $\tau_0$, which is asymptotically equal to the lag-1 autocorrelation $|r(1)|$ for a binary process. Here $\rho_{11}$ and $\rho_{01}$ are, respectively, the conditional probabilities of the occurrence of rain, given that rain occurred or did not occur on the previous day. The degrees of freedom used in estimating noise in (6) are given by

$$\{[(T(1 - d)/(1 + d)] - 1\} J, \quad T = 62 \quad \text{and} \quad J = 19 \text{ here}.$$

The above formula is valid for the first-order Markov Chain process. Klugman (1983) has given the procedures for other orders of the Markov Chain. However, the results may not be very sensitive to the order of the Markov Chain, and the effect of considering the Markov process of order one as against some other true order may not be as serious as ignoring the tem-
poral dependence in rainfall events altogether. Further, the studies conducted by the authors (Singh et al., 1981, and some later studies) have shown that the Markov Chain of order one fits the monsoon rainfall data satisfactorily.

4. Results

For the given degrees of freedom, an $F$-value of two is significant at $\leq 1\%$ level. The $F$-value of two also signifies that the signal-to-noise ratio (i.e., $F - 1$) is equal to one and that half of the observed interannual variability is potentially predictable. The higher the value of $F$, the more predictable the field. However, the null hypothesis of no signal should not be accepted for the cases of $F \leq 2$ since the confidence limits on $F$ are quite large (Trenberth, 1985; Hayashi, 1982).

a. 700-mb height field

The results for the 700-mb height fields are presented in Fig. 1. The most important feature of the distribution of $F$-values (Fig. 1a) is the general decrease with the increase in the latitude. The values are very high, exceeding eight at places south of $15^\circ$N latitude. There is also a very sharp gradient between $15^\circ$ and $20^\circ$N latitude. The area north of $20^\circ$N can be divided into three longitudinal sectors. Over the middle sector, of about $10^\circ$ longitude width, the value of $F$ exceeds two, and the $F$-values are less than two over the other two sectors. Northeast of the head of the Bay, the values of $F$ are even less than one, meaning that the field over this region is unpredictable. This pattern suggests that

the variation in fields near or south of $15^\circ$N have a very important bearing on the predictability of the monsoon circulation. This behavior of the $F$-statistic can be partly understood from the distribution of $\tau_0$ (Fig. 1b). As can be seen, the $\tau_0$ is minimum ($\leq 2$ days) between latitudes $10^\circ - 15^\circ$N and maximum ($\geq 5$) near $25^\circ$N latitude.

It may be noted that the main storm/synoptic activity during July-August lies between $20^\circ$ to $25^\circ$N latitude, where the $\tau_0$ value is higher. This is contrary to the results obtained for the extratropical regions, where smallest $\tau_0$ are found in the regions of main storm tracks in both the hemispheres (Trenberth, 1985). This may be due to the longer time scales, viz., quasi-bi-weekly and 40-50 day scales, noted in the monsoon activity. Although the lifetime of individual storms is a few days, the monsoon remains active or weak for longer periods. The lag-1 autocorrelation $[\rho(1)]$ is significantly different from zero ($\geq 0.5$) over the entire region (Fig. 1c), with values exceeding 0.7 between $15^\circ$ and $25^\circ$N.

The values of $\tau_0$ could be determined after finding the appropriate order of the autoregression (AR) models. This, however, has not been attempted here. We have computed $\tau_0$ appropriate for AR(1) and AR(2) models (not presented); this required estimates of $\rho(1)$ and $\rho(2)$. On comparison, the $\tau_0$ values computed for AR(2) model were found to agree closely with the values computed by using (5). Trenberth (1985) found that generally AR(2) is the proper model for the geopotential heights over the Australasian region; the model exhibits quasi-periodicities in a 10-50 day range. The summer monsoon circulation over the Indian re-

![Fig. 1](image_url)

Fig. 1. (a) $F$-value, i.e., signal-to-noise ratio; (b) $\tau_0$—relaxation time; (c) lag-1 autocorrelation coefficient; (d) lag-3 autocorrelation coefficient for 700 mb contour height fields.
gion exhibits a prominent quasi-periodicity in the 40-50 day range, which supports that the AR(2) may be the best model for the circulation fields over this region also. It may be reiterated that, as pointed out by Trenberth (1984a), the method followed for estimation of the autocorrelations ($r$'s) and $\tau_0$ in the present study will produce negatively biased estimates. The different methods of estimating $r$ and the different AR models can give slightly different results. (See Trenberth, 1984a,b.) The analysis of the impact of various methods on estimation of $r$ and $\tau_0$ is postponed till a future study.

The above computations were also repeated to examine the signal-to-noise ratio of seasonal (June to September) contour heights and contour height anomalies. The main results for seasonal contour height anomalies were similar to that for July-August but showed higher autocorrelations and $\tau_0$ values due to the presence of seasonal cycle.

b. Mean sea level pressure anomaly fields

The results for mean sea level pressure anomaly fields are presented in Fig. 2. Over major parts of the country, the values of $F$ are less than two, with a minimum near the head of the Bay of Bengal. $F$-values are more than two south of 15°N and over Northwest India. The latter area is on the eastward edge of a large region in the subtropics where a high signal-to-noise ratio occurs. (See Fig. 3 of Madden, 1983.) A comparison of the spatial patterns of the $F$-ratio for 700-mb contour height and mean sea level pressure anomaly (Figs. 1 and 2) reveals that the $F$-values differ near 15°N, and secondly, there is a minimum in the $\tau_0$ field for 700-mb contour height fields in the south Bay of Bengal near 10°N latitude. This is mainly due to the differences in autocorrelations at lags greater than one, as seen by the lag-5 autocorrelations which result in different $\tau_0$ patterns. For the mean sea level pressure anomaly, the autocorrelation falls off more slowly over and near the head of the Bay of Bengal and relatively faster over the remaining parts of the country. The presence of the minimum in $\tau_0$ over the south of the Bay of Bengal in the 700-mb contour height fields can be partially accounted for by the greater day-to-day variability of the 700-mb circulation associated with the frequent appearance of a near-equatorial oceanic intertropical convergence zone (ITCZ) which is most prominent at the 700-mb level in these longitudes (Sikka et al., 1981).

c. Rainfall fields

Although we have computed the signal-to-noise ratio and the rainfall amounts for the rainy days, we present here the results for only the rainfall amounts, as the results for the rainy days are not very promising. The computations were done for three thresholds for defining a rainy day, and these results were similar. Hence we present (Fig. 3) the results for a 2.5 mm threshold, which is the official definition of the rainy day in India.

The values of $F$ exceed three over a contiguous region consisting of East M.P., Orissa, Bihar and Bengal lying to the east of 80°E longitude and exceed two over the west coast and the surrounding regions. Thus, over large parts of India, more than 50% of the interannual variability is potentially predictable. There is, indeed,

![Fig. 2. As in Fig. 1 but for mean sea level pressure anomaly field.](image-url)
ample evidence that a few regional or remote parameters are able to explain more than 50% of the variance of the monsoon rainfall, and the results in Fig. 3 support this conclusion. It may be noted that the temporal dependence in the daily rainfall is more for the subdivisions lying in the western parts of the country (Singh et al., 1981); hence, this part of the country shows the higher value of $\tau_0$. Statistical models developed for predicting five-day and weekly rainfall show higher forecast skills for this region. The daily variance being comparable, the higher $\tau_0$ in western parts increases the estimates of noise and, hence, decreases the potential predictability. The geographical distribution of the $F$-statistic is thus consistent with the above concept.

5. Conclusions and future plans

On the basis of the above results the following conclusions are drawn:

(i) The potential predictability of seasonal lower-tropospheric atmospheric fields is less over the area normally occupied by the monsoon trough, and it generally increases with decreasing latitude. The potential predictability is higher over northwest India and adjoining regions.

(ii) About half of the interannual variability of rainfall is predictable over the entire country. For some regions the predictable variance is more than 50%. The total rainfall amounts are more predictable than the number of rainy days.

There exists further need for studying the predictability of the tropical atmosphere and rainfall. The daily station data of mean sea level pressure, upper-air wind data and rainfall over a larger region, e.g., covering the Indo-Pacific Southern Oscillation region or the entire tropical region, can be used to analyze the potential predictability of these fields. The satellite-observed cloudiness and outgoing longwave radiation data now available for over 10 years can enable us to study the potential predictability of the entire tropical features. These data may help us to understand the longitudinal asymmetries, if any, in potential predictability. Certain modes of the atmosphere may be more predictable than others. To examine this, predictability studies should be conducted on empirical orthogonal functions of the circulation and rainfall fields. Studies are also required on the predictability of mean features in the other seasons. To partially fulfill this aim, the daily 700-mb contour height data over an extended region for a 15-year period and daily rainfall data of 380 stations over India for an 80-year period have been obtained. These data are, at present, being processed. The empirical orthogonal functions of the daily circulation and rainfall fields have been determined. It is hoped that an analysis of these new data, with slight modification of the methodology, may add to our current knowledge about the potential predictability of the circulation and rainfall during the monsoon and other seasons.

After this paper was completed and while it was being revised, recent work of Trenberth (see references), pointing out the pitfalls in the estimation of autocorrelation and the confidence interval of $F$-values in the face of a signal, has come to our attention, as has the work of Thibaux and Zwiers (1984). We intend to examine the various methodologies in a future study to be conducted on the enlarged data mentioned above.

Acknowledgments. The authors are thankful to Dr. K. E. Trenberth for supplying the copies of some of the galley proofs and reprints which substantially improved our understanding of the subject and, possibly, the manuscript. We are also thankful to the reviewers for constructive suggestions which led to improvement in the quality of the paper. We are grateful to the Director, Indian Institute of Tropical Meteorology, for the facilities provided, to Shri D. R. Sikka, F.A.Sc., Head, Forecasting Research Division, for encouragement and to the India Meteorological Department for supplying the rainfall data. Thanks are due to the several colleagues and associates who helped in a variety of ways to prepare the 700-mb contour height data and to Shri D. K. Paul for supplying the mean sea level pressure anomaly data. Thanks are also extended to Miss Chabi P. Ghosh and Shri K. D. Barne, for typing the manuscript.

REFERENCES


