Extreme Daily Rainfall Events and Their Impact on Ensemble Forecasts of the Indian Monsoon

D. B. STEPHENSON, K. RUPA KUMAR, F. J. DOBLAS-REYES, J.-F. ROYER, AND F. CHAUVIN

Météo-France, Toulouse, France

S. PEZZULLI

*Dipartimento di Studi Geoeconomici, Statistici, Storici per l'Analisi Regionale, Universitá di Roma, ''La Sapienza,''
Rome, Italy*

ABSTRACT

The Indian summer monsoon rainfall is the net result of an ensemble of synoptic disturbances, many of which are extremely intense. Sporadic systems often bring extreme amounts of rain over only a few days, which can have sizable impacts on the estimated seasonal mean rainfall. The statistics of these outlier events are presented both for observed and model-simulated daily rainfall for the summers of 1986 to 1989. The extreme events cause the wet-day probability distribution of daily rainfall to be far from Gaussian, especially along the coastal regions of eastern and northwestern India. The gamma and Weibull distributions provide good fits to the wet-day rainfall distribution, whereas the lognormal distribution is too skewed. The impact of extreme events on estimates of space and time averages can be reduced by nonlinearly transforming the daily rainfall amounts. The square root transformation is shown to improve the predictability of ensemble forecasts of the mean Indian rainfall for June 1986–89.

1. Introduction

Monthly and seasonal mean rainfall amounts, although often used to describe the general behavior of the Asian summer monsoon, can give the misleading impression that the monsoon is a robust slowly evolving system. The mean rainfall, however, is the rainfall averaged over many sporadic weather events having spatial scales from 100 to 1000 km. For example, on 25 August 1997, the *Times of India* reported that:

> At least 23 people including four children and two women have died and thousands rendered homeless all over the state in the heavy rains that continued to batter Mumbai [Bombay] and other parts of Maharashtra for the third consecutive day on Sunday. The city, which recorded the highest daily rainfall in six years—407.6 mm in Colaba and 445.6 in Santa Cruz till 8:30 am on Sunday morning was gradually limping back to normalcy after the utter chaos witnessed during the last two days with the breakdown in the rail, road and air network . . . .

Typical June–September wet-day average rainfall at

Mumbai (Bombay) is around 18 mm day^{-1} with a daily standard deviation of 28 mm day^{-1}; hence, more than 400 mm of rain in one day represents an *extreme event* or *outlier* more than 10 standard deviations greater than the mean. Outliers are not uncommon in Indian monsoon rainfall and are due to convection mainly associated with monsoon depressions (Sikka 1977) and midtropospheric cyclones (Keshavamurthy 1973). Seasonal mean monsoon rainfall is a statistical average over many individual weather events and can only be perfectly forecast by knowing the precise future evolution of all the events. However, for nonlinear systems such as the atmosphere, initial perturbations (small uncertainties) can diverge exponentially fast, thereby imposing a fundamental limit on our ability to predict weather events beyond a certain time horizon (Lorenz 1963). A measure of this horizon is given by the time needed for initial perturbations to double in amplitude, which for midlatitude geopotential height appears to be of the order of 1–2 days and for monsoon rainfall could be even less. The inability to perfectly predict the synoptic evolution throughout the whole season has led some authors to conclude pessimistically that prediction of climatic features, not directly tied to the annual march of the seasons, is doomed to fail (Ramage 1971). Nevertheless, more slowly evolving environmental factors, such as land and sea surface temperatures and planetary-scale

Corresponding author address: Dr. D. B. Stephenson, Laboratoire de Statistique et Probabilitiés, Université Paul Sabatier, 118, Route de Narbonne, F-31062 Toulouse Cedex, France.
E-mail: stephen@cict.fr

atmospheric circulations, can influence the formation and development of rain-bearing systems (Charney and Shukla 1981). These factors often lead to predictable *signals* that can be projected out by compiling statistics over all the weather events (e.g., seasonal rainfall totals), and it is this hope that underlies current attempts to produce seasonal rainfall forecasts. Statistical estimates, however, contain unpredictable *sampling noise* caused by natural weather fluctuations, which are sometimes misleadingly referred to as ''chaos'' (Palmer 1994). For example, a single unpredictable extreme daily weather event at Mumbai can easily contribute 400 mm to the total seasonal rainfall and, thereby, increase the seasonal mean rainfall by more than 3 mm day^{-1}. When one compares this with the interannual standard deviation of the mean June–September Mumbai rainfall of 4 mm day^{-1}, one realizes that isolating the predictable signal from the noise may be a rather delicate problem. Rodwell (1997) also cited a case in which a single strong monsoon depression over northwest India on 12–14 July 1994 could easily be identified in the seasonal mean all-India rainfall anomaly. Such extreme events act as a source of severe sampling noise that can easily spoil sample estimates of seasonal mean rainfall.

This study investigates extreme rainfall events and the impact that they have on ensemble forecasts of Indian monsoon rainfall. Section 2 of this article discusses the observed and model-generated daily rainfall data used in the study. Section 3 summarizes the rainfall behavior over India and shows that extreme rainfall events are more prevalent on the east and northwest coasts of India. The probability distribution of the wet-day daily rainfall is briefly discussed in section 4, where it is demonstrated that good fits are provided by the gamma and Weibull distributions but not by the log-normal distribution. Section 5 reviews the use of non-linear transformations as a way of reducing the contribution from extreme events, and section 6 then illustrates the impact of the square root transformation on the evolution of a generalized all-India rainfall index. Section 7 demonstrates how the square root transformation can improve ensemble forecasts of Indian rainfall for June 1986–89. Section 8 concludes the article with a brief summary.

2. Data sources and treatment

This study focuses on Indian monsoon daily rainfall data over the short period of 488 days covering June–September 1986–1989. India received spatially averaged rainfall totals of 743 mm (1986 normal monsoon), 697 mm (1987 weak monsoon), 962 mm (1988 strong monsoon), and 867 mm (1989 normal monsoon). These summers have been the focus for the European monsoon project SHIVA (Studies of the Hydrology, Influence and

Variability of the Asian Summer Monsoon).[1] The short-period time slice allows us to focus in detail on the daily rainfall, yet is long enough to justify the conclusions made in this preliminary study. More reliable estimates may be obtained by considering longer periods, but at the risk of encountering complications due to epochal nonstationarities and/or measurement inhomogeneities (Pant et al. 1988).

a. Observed Indian daily rainfall amounts 1986–89

This study uses a dense network of about 400 rain-measuring stations (gauges) spread homogeneously throughout the nonmountainous part of India (Fig. 1). Not all the station reports are available each day, and from June to September 1986–89 only 212 stations reported rainfall totals on more than 75% of the days. Only the quality-controlled time series from the subset of 212 stations obtained from the archives at the India Meteorological Department (IMD, Pune) have been used to make the statistics. The statistics have then been gridded onto a 1° resolution grid by using an inverse square distance kernel with a search radius limited to 500 km, similar to the method proposed by Cressman (1959). Kriging was also tested and gave similar results.

b. Model-simulated rainfall 1986–89

To assess how a state-of-the-art climate model simulates daily extremes in monsoon rainfall, we also present results based on gridded daily rainfall amounts produced by the Météo-France general circulation model: ARPEGE. General circulation models capture some of the broad-scale features of the Asian summer monsoon yet are generally poor at simulating regional-scale features such as the monsoon over India (Stephenson et al. 1998a). Nevertheless, a preliminary comparison has been made of the model rainfall statistics with those obtained from Indian station data. Unfortunately, because of the lack of rain gauge measurements and the overly large errors in satellite-based daily estimates, comparison is not possible over the Indian Ocean where it rains the most.

ARPEGE is a state-of-the-art 19-level spectral climate model developed from the weather forecasting model operationally employed at Météo-France and the European Centre for Medium-Range Weather Forecasts (ECMWF). Deep subgridscale cumulus convection is parameterized using a Kuo-type scheme that requires both low-level large-scale moisture convergence and a conditionally unstable moist-adiabatic temperature profile for convection to occur. Shallow convection is parameterized using a modified Richardson number

---

[1] The SHIVA project is described in detail online at http://www.met.rdg.ac.uk/shiva/shiva.html.
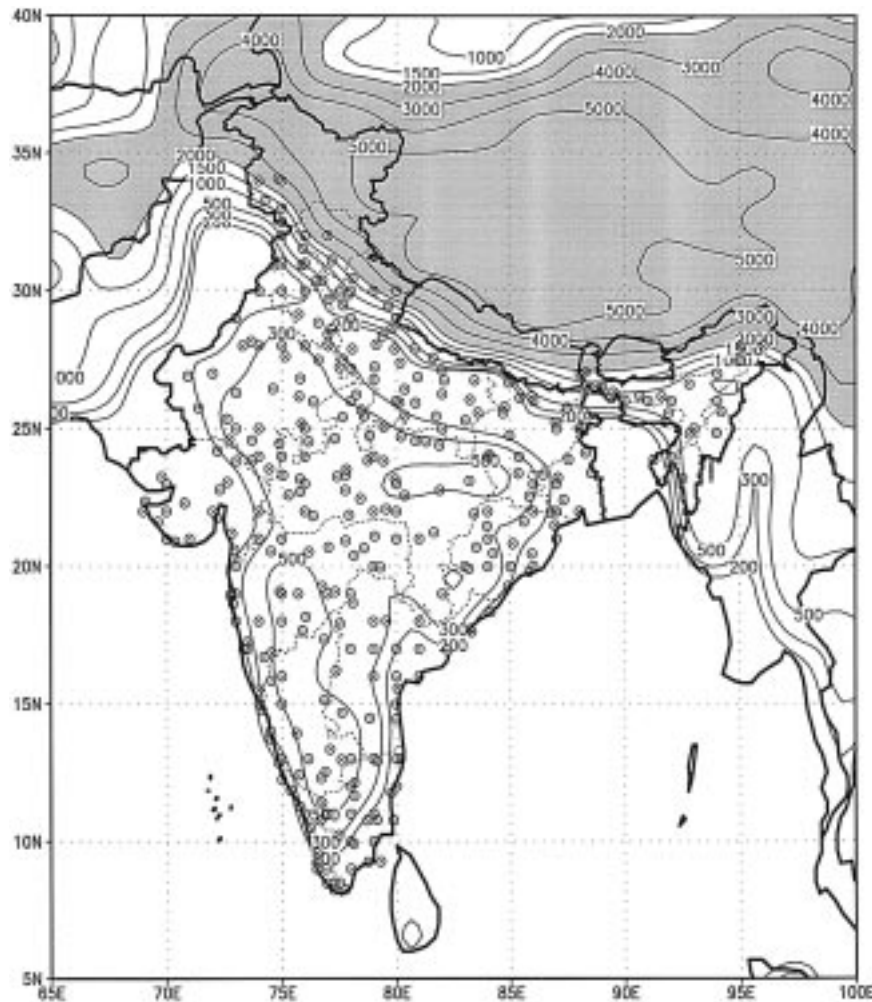
FIG. 1. Location of the rain gauge stations used in the present study. Altitude is indicated by contour lines marked in meters above sea level.

scheme, and stratiform cloudiness is calculated by comparing the humidity profile to that of a critical profile. Surface land temperature, deep soil temperatures, and surface water content are calculated prognostically using an interactive soil–biosphere scheme. The dependence of the mean summer monsoon on horizontal resolution has been investigated by performing simulations spectrally truncated at 21, 31, 42, and 63 zonal wavenumbers. This study will only present statistics for the most realistic monsoon simulation made at the medium horizontal resolution of 300 km (T42 truncation). A full description of the model and the monsoon results can be found in Stephenson et al. (1998a).

c. Wet and dry days

Precipitation is a discontinuous process consisting of both *wet* and *dry* days. Daily rainfall amounts are described by the mixed distribution $p(x) = p(x|\text{dry})p_{\text{dry}} + p(x|\text{wet})p_{\text{wet}}$, where x is the daily rainfall total and $p_{\text{wet}} = 1 - p_{\text{dry}}$ is the probability of having a wet day. A number of difficulties can be avoided by focusing on rainfall amounts on only the wet days. For example, the normal onset date of the monsoon rainfall varies between the start of June in the south of India to the start of July in the northwest of India and hence northwest India generally has less wet days in June than does southern India. Including both wet and dry days leads to northwest India having a smaller mean and a larger skewness than southern India. Such biases are undesirable and therefore only the rainfall on wet days have been used to calculate the statistics presented in this study. Since rainfall totals are recorded at IMD in tenths of millimeters, 0.1 mm represents an obvious cutoff for defining wet days and has been adopted in this study both for the station data and for the model-generated gridpoint data. Other definitions are possible, such as *rainy days* being those receiving more than 2.5 mm (0.1 in.) of rainfall per day (Soman and Krishna Kumar 1990). Statistics characterizing extreme behavior mea-
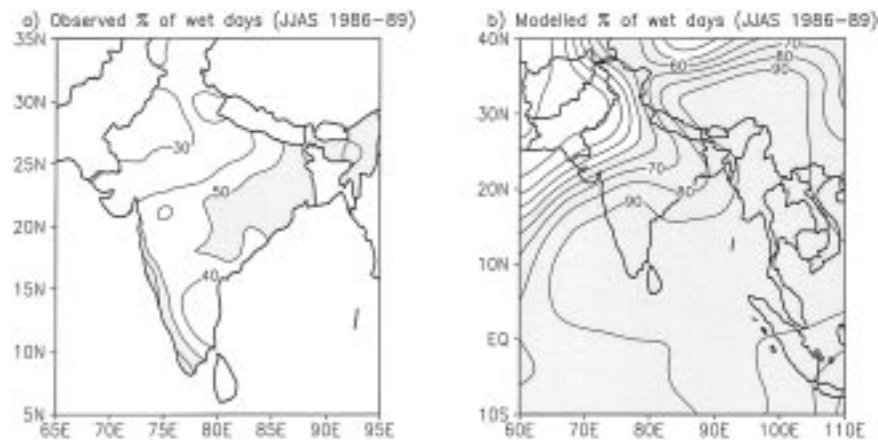
FIG. 2. Frequency of occurrence of wet days for (a) the Indian observations, and (b) the rainfall simulated by the model over the Asian region, estimated over the June–September summers 1986–89. A wet day is defined as a day receiving more than 0.1 mm of total rainfall.

sure the large rainfall tail of the probability distribution and are, therefore, relatively insensitive to the choice of the small nonzero wet-day rainfall cutoff. A different threshold of 1 mm day^{-1} was found to give very similar results to those presented here. In this preliminary investigation, the total number of wet days at a particular station or grid point in India is typically between 100 and 400. Figure 2a shows the frequency of occurrence of wet days observed over India for June–September 1986–89. Typically 40%–60% of the days are *wet* with less wet days occurring over the northwestern desert region of India. The model gives a similar northwest–southeast gradient (Fig. 2b) yet has a much larger fraction of wet days over most parts of India. This is explained by the model's convection scheme having a tendency to drizzle, and also gridded model rainfall data being an average over a grid box rather than being point-like station data (Skelly and Henderson-Sellers 1996).

3. Geographical variations in wet-day rainfall characteristics

Table 1 presents statistics for selected stations in India calculated using rainfall observations on wet days in June–September 1986–89. Similar statistics have been calculated at all the quality-controlled stations in India and have been gridded to produce the maps in Fig. 3. The mean wet-day rainfall rate is 10–20 mm day^{-1} with two distinct maxima located along the southwest coast and over northeast India (Fig 3a). Since standard deviations are somewhat larger around 15–30 mm day^{-1} yet exhibit a similar spatial distribution to that of the mean, it is more instructive to present the ratio of the standard deviation to the mean, referred to as the *coefficient of variation* (Fig. 3b). The coefficient of variation (CV) has typical values of 1.2–2.0 with largest values occurring in the highly variable east coast region of Andra Pradesh and the west coast region around Gu-

TABLE 1. Observed wet-day statistics for a selection of Indian stations. The arithmetic mean (in mm day^{-1}), standard deviation (std dev in mm day^{-1}), coefficient of variation (CV), skewness, and kurtosis statistics are calculated for the wet days having more than 0.1 mm of total rainfall. Max gives the maximum daily rainfall totals (in mm) recorded during this period, and NWD gives the sample size, that is, the number of recorded wet days out of the total of 488 days in June–September 1986–89.

| Station	Location	Mean	Std dev	CV	$\sqrt{b_1}$	$b_2 - 3$	Max	NWD
Jalpaiguri	26.5°N, 88.7°E	36.4	45.3	1.24	2.3	6.6	288	303
Goa	15.5°N, 73.9°E	24.2	32.6	1.35	3.4	21.3	335	402
Trivandrum	8.50°N, 77.0°E	11.2	15.9	1.41	2.7	9.6	119	319
Nagpur	21.1°N, 79.1°E	13.4	19.8	1.48	2.4	6.2	107	249
Mumbai	18.9°N, 72.8°E	18.5	28.2	1.52	2.5	6.7	183	395
Tezpur	26.6°N, 92.8°E	15.9	24.3	1.52	3.5	16.6	185	278
Calcutta	22.5°N, 88.3°E	17.1	26.5	1.55	4.0	25.2	259	341
Jammu	32.5°N, 75.0°E	22.9	36.3	1.59	3.2	14.2	270	186
Hyderabad	17.5°N, 78.5°E	11.1	17.9	1.61	3.3	14.9	141	238
New Delhi	28.6°N, 77.2°E	11.7	20.0	1.71	3.7	19.9	159	150
Pune	18.5°N, 73.9°E	8.2	14.6	1.78	4.3	27.7	142	287
Visakhapatnam	17.7°N, 83.2°E	13.4	25.3	1.89	3.3	12.7	178	205
Kakinada	16.9°N, 82.2°E	10.6	23.3	2.20	8.2	91.3	296	247
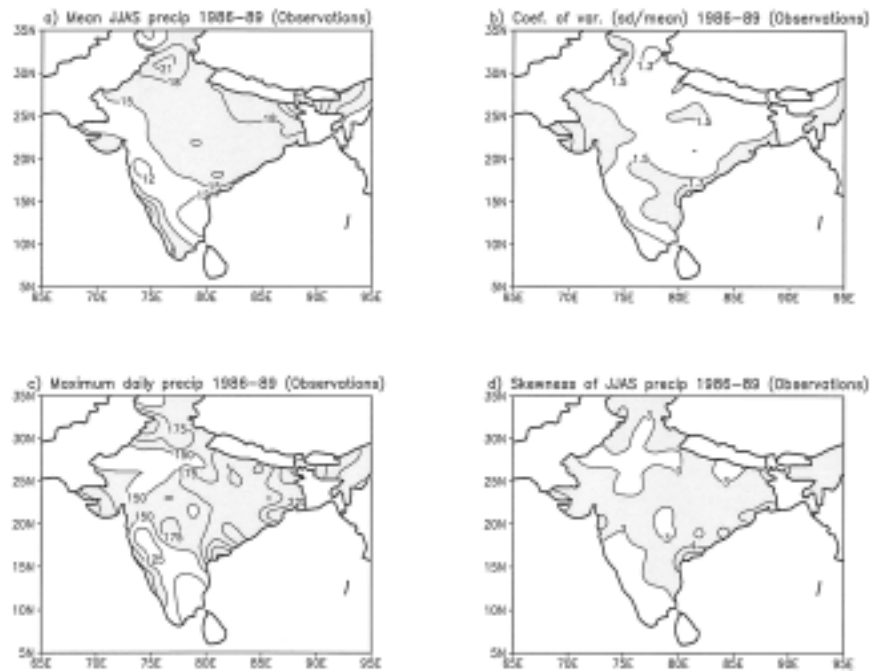
FIG. 3. Statistics of observed wet-day daily rainfall estimated over the June–September summers 1986–89 (488 days): (a) mean (mm day^{-1}), (b) coefficient of variation (standard deviation/mean), (c) maximum daily rainfall amount (mm day^{-1}), and (d) moment measure of skewness ($\sqrt{b_1}$). Statistics were calculated for each of the Indian stations having available rainfall records and then were interpolated onto a 1° latitude–longitude grid before plotting (refer to section 2a for details of interpolation).

jarat. These transition regions lie to the south and north, respectively, of regions receiving a large amount of rain-bearing systems, and occasionally receive strong synoptic systems. Such intermittency leads to low values of mean rainfall while standard deviations remain typical and, thereby, give rise to larger values of CV. Typical maximum daily rainfall totals recorded over the June–September periods of 1986–89 are around 100–250 mm day^{-1} and have largest values in the aforementioned coastal regions (Fig. 3c). The east coast maxima are often associated with monsoon depressions that develop over the Bay of Bengal, whereas the west coast maxima are mainly due to cyclones and troughs originating over the Arabian Sea. The maximum daily values are roughly 10 standard deviations larger than the typical mean values and, therefore, would have a 1 in 5 × 10^{20} chance of occurring if the rainfall were Gaussian distributed!

The deviation of a probability distribution from normality (Gaussian) can be quantitatively assessed by estimating the *skewness* (asymmetry) $g_1 = \sqrt{b_1} = m_3 m_2^{-3/2}$, and *kurtosis* (flatness) $g_2 = b_2 - 3 = m_4 m_2^{-2} - 3$, where $m_k = E(x - \overline{x})^k$ is the estimate of the kth-order moment about the sample mean (Kendall et al. 1983). These statistics quantify intermittency and extreme behavior and are zero for Gaussian-distributed variables. Skewness $\sqrt{b_1}$ has been calculated for all the Indian stations and is presented in Fig. 3d and for se-

lected stations in Table 1. Values typically exceed 3 over much of India, with largest values occurring on the east coast near Kakinada (16.9°N, 82.2°E) and along the northwest coast north of Mumbai—regions where the coefficient of variation was also found to be largest. Assuming 100 degrees of freedom, there is less than 1% chance that a Gaussian distribution would give $\sqrt{b_1}$ greater than 0.75 (Pearson and Hartley 1972). The spatial distribution of kurtosis was also found to qualitatively resemble that of skewness (Stephenson et al. 1998b). An alternative more robust measure of skewness is given by the difference between the mean and the mode/median (Shea and Sontakke 1995). Karl Pearson considered the dimensionless measure of skewness given by the difference between the mean and the mode divided by the standard deviation (Kendall et al. 1983). For rainfall amounts, where the mode is close to zero, Pearson's skewness becomes approximately equal to the *inverse* of the coefficient of variation, in apparent contradiction with the moment measure of skewness $\sqrt{b_1}$, which is strongly *positively* correlated with CV in Table 1 and Fig. 3. The paradox is resolved by noting that the rainfall distribution does not have a vanishing gradient at its mode and, therefore, does not belong to the well-behaved Pearson class of distributions (Kendall et al. 1983). The Pearson measure of skewness can therefore be very misleading when applied to rainfall amounts and should be avoided in such studies.
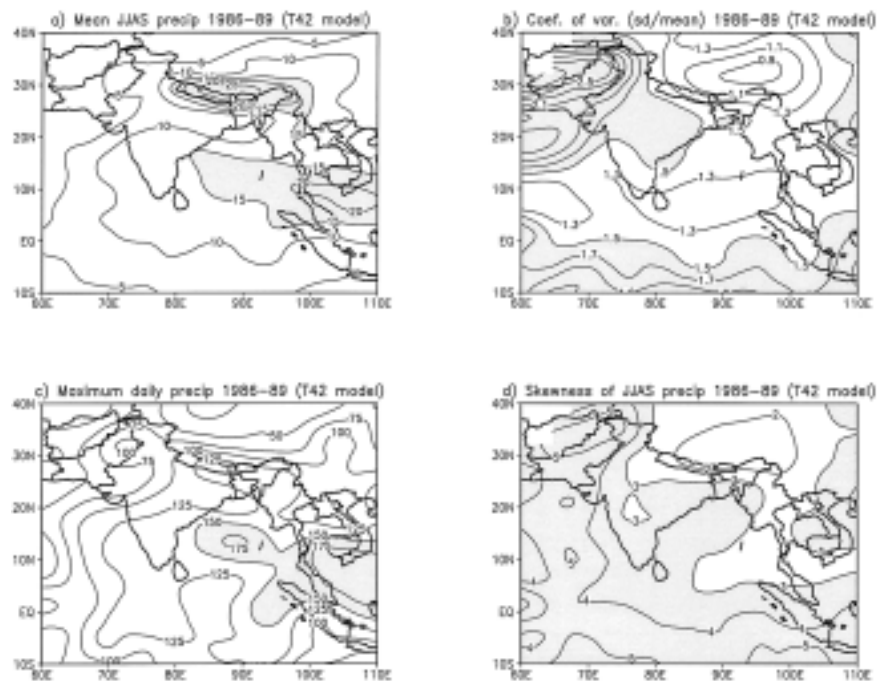
FIG. 4. Statistics of modeled wet-day daily rainfall estimated over the June–September summers 1986–89 (488 days): (a) mean (mm day^{-1}), (b) coefficient of variation (standard deviation/mean), (c) maximum daily rainfall amount (mm day^{-1}), and (d) moment measure of skewness ($\sqrt{b_1}$). Statistics were calculated for each of the model grid points spaced roughly every 2.8°. Note that contours are every 5 mm day^{-1} in (a) compared to every 3 mm day^{-1} for the observations in Fig. 3a.

Figure 4 shows similar wet-day statistics calculated for the model-generated daily gridded rainfall amounts. There is excessive mean rainfall over the southern slopes of the eastern Himalaya, and the mean rainfall maximum over India is displaced too far south (Fig. 4a). These are common problems often encountered in model simulations of the monsoon and are discussed in Stephenson et al. (1998a). The coefficient of variation has values close to 1.5 similar to those observed yet fails to capture the local maxima along the Indian coasts that were noted in the observed skewness (Fig. 4b). In Fig. 4c, the maximum daily rainfall amounts are around 100–175 mm day^{-1} and are, therefore, slightly smaller than those noted in the observations over India. Whereas standard deviations stay approximately constant, maximum daily values were found to increase with increasing model horizontal resolution from values around 100 mm day^{-1} at 600-km resolution (T21 truncation) up to more realistic values around 200 mm day^{-1} at 200-km resolution (T63 truncation). East coast extreme events are often caused by monsoon depressions having spatial scales of around 500 km (Krishnamurti et al. 1975), which are now beginning to be resolved by higher-resolution atmospheric models (Lal et al. 1995). Although an advantage for weather prediction, the existence of more extreme events could lead to undesirable noisiness in climate statistics generated by future models having higher horizontal resolutions. The model gives com-

parable values of skewness to those observed with a maximum of 3.5 over the head of the Bay of Bengal (Fig. 4d).

4. Probability distribution of wet-day rainfall amounts

The probability distribution of monthly mean Indian rainfall was first discussed in the studies by Sankaranarayanan (1933) and Pramanik and Jagannathan (1953). In a later study, Mooley and Crutcher (1968) demonstrated that the gamma distribution provides good fits to monthly mean rainfall totals observed at numerous Indian stations. The gamma distribution has been extensively used to fit rainfall amounts in other studies (Barger and Thom 1949; Thom 1958; Ison et al. 1971; Shenton and Bowman 1973). For gamma-distributed rainfall amounts, the probability dF of an amount occurring between x and $x + dx$ is given by

$$dF = \frac{a^\gamma}{\Gamma(\gamma)} e^{-ax} x^{\gamma-1} \, dx, \qquad (1)$$

where γ is the shape parameter, a is the inverse scale parameter, and $\Gamma(\gamma)$ is the gamma function given in mathematical tables. The distribution has a mean of γ/a, a standard deviation of $\sqrt{\gamma}/a$, a coefficient of variation of $1/\sqrt{\gamma}$, a skewness $\sqrt{b_1} = 2/\sqrt{\gamma}$ (twice the coefficient of variation), and a kurtosis $b_2 - 3 = 6/\gamma$. In the Gauss-
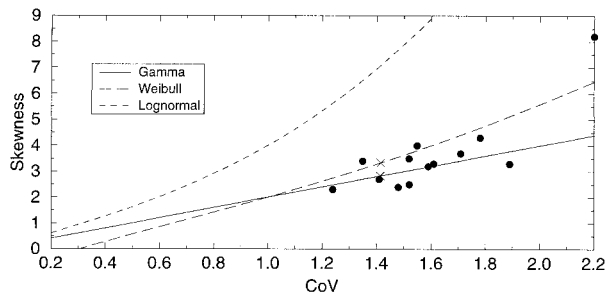
FIG. 5. Scatterplot of the skewness vs the coefficient of variation for the 13 Indian stations given in Table 1 (filled circles). Also shown are the functional relationships expected for the gamma distribution (solid line), the Weibull distribution (long dashes), and the lognormal distribution (short dashes). The $\times$ signs mark points with $\gamma = 0.5$ for the gamma distribution and $\alpha = 0.72$ for the Weibull distribution.

ian limit as $\gamma \rightarrow \infty$, the skewness and kurtosis tend to zero. For $\gamma < 1$, the coefficient of variation exceeds unity and the distribution is strongly positively skewed with the mode at zero (an inverted J-shape distribution).

The shape parameter γ can be crudely estimated by equating the coefficient of variation to $1/\sqrt{\gamma}$ (method of moments). The CV for daily wet-day Indian rainfall is typically about 1.5 giving an estimate of $\gamma \approx 0.5$. A gamma distribution with $\gamma = 0.5$ has skewness $\sqrt{b_1} = 3$ and kurtosis $b_2 - 3 = 12$, both of which are consistent with the observed values in Table 1. Larger values of skewness and kurtosis occur at certain locations (e.g., Kakinada), where maximum likelihood estimates of γ are found to be slightly less than 0.5 (Stephenson et al. 1998b). Figure 5 shows a scatterplot of the skewness and CV values for the stations in Table 1. The cloud of points lie close to the $\sqrt{b_1} = 2$CV linear relationship expected for a gamma distribution. Figure 5 also includes curves for two other distributions that have been applied to rainfall totals: the lognormal distribution (Kedem et al. 1990; Kedem et al. 1994; Wilheit et al. 1991), and the Weibull distribution (Weibull 1951; Essenwanger 1986; Wong 1977; Wilks 1989). The Weibull distribution also fits the station data and has the advantage of having a simple cumulative distribution $F(x) = 1 - e^{-(ax)^{\alpha}}$, where α is a shape parameter (Stephenson et al. 1998b). The lognormal distribution, which assumes that $\log x$ is Gaussian distributed, is too skewed (and leptokurtic) to fit the station observations (Fig. 5). This finding is in apparent contradiction to that of Kedem et al. (1990) and Kedem et al. (1994) who found that the lognormal distribution gave better χ^2 fits than did the gamma distribution. In their studies, however, more skewed instantaneous rainfall rates were being examined rather than daily totals. Nevertheless, a χ^2 least squares approach is inappropriate for fitting or judging highly skewed distributions that necessitate instead the use of maximum likelihood and Kolmogorov–Smirnov methods (Kendall et al. 1983). Other possible distributions are concisely reviewed in Öztürk (1981).

The gamma distribution is routinely used to provide confidence limits for forecasts of monthly and seasonal mean rainfall amounts (Ropelewski and Jalickee 1983; Ropelewski et al. 1985). Numerical integrations are invariably required to obtain such confidence limits. However, for the special case when $\gamma = 0.5$, the probability of a wet day having more than X millimeters of rainfall is simply given by $P(x > X) = 2\Phi(z > X^{1/2}m^{-1/2})$, where m is the mean rainfall amount and $\Phi(z > Z)$ is the area under the standard normal curve for values of $z > Z$ (Stephenson et al. 1998b). For example, this implies that there is approximately only 10% chance of wet-day Indian rainfall daily totals exceeding four times their mean values. Finally, for independent gamma-distributed data, the arithmetic mean over n days (or n stations) is also gamma distributed, yet with a larger shape parameter of $n\gamma$ (Kendall et al. 1983). Such an increase in the shape parameter for longer time averages can be noted in the gamma values for Mumbai: 0.5 for daily totals (this study), 4.9 for monthly totals (Mooley 1973a,b), and 13.6 for seasonal 4-month total (Mooley and Appa Rao 1971). It might be possible to exploit this scaling property of the gamma distribution to extrapolate climate forecasts of rainfall down to regional or point values, for example, for downscaling studies similar to that in Osborn and Hulme (1997).

5. Reducing the impact of extreme events

Spatial maps of daily monsoon rainfall often reveal conspicuous small-scale bull's-eye patterns associated with extreme events. These noisy features are often deemphasized in rainfall maps by choosing unevenly spaced contours such as 1, 2, 4, 8, etc. millimeters per day. This is equivalent to contouring nonlinearly transformed rainfall totals using evenly spaced contours. The resulting maps are less noisy and more clearly reveal larger-scale features. The use of nonlinear transformations to reduce outlier noise not only improves the visual quality of contour maps but can also be used to improve the robustness of statistical analyses (Dolby 1963). Reliable statistical estimation often depends on the sampled data being close to Gaussian, thereby ensuring the relatively rare occurrence of troublesome outliers. Monotonic nonlinear transformations are often applied by statisticians to make data more Gaussian before being analyzed. Katz (1983) concluded his study of procedures for making inferences about precipitation changes by stating that, "transformations other than the logarithm could be applied to precipitation intensities in an attempt to obtain data having a more nearly Gaussian distribution." In particular, the class of power law transformations, $y = (x + \lambda)^{\nu}$, have been widely used by statisticians (Haldane 1938; Box and Cox 1964, and many others). The parameter ν determines the *strength* of the transformation and is typically chosen to be between zero and unity (Tukey 1955). The logarithmic transformation $y = \log(x + \lambda)$ is obtained as the special

a) Daily observed rainfall amounts at Colaba Observatory, Mumbai (1986–89)

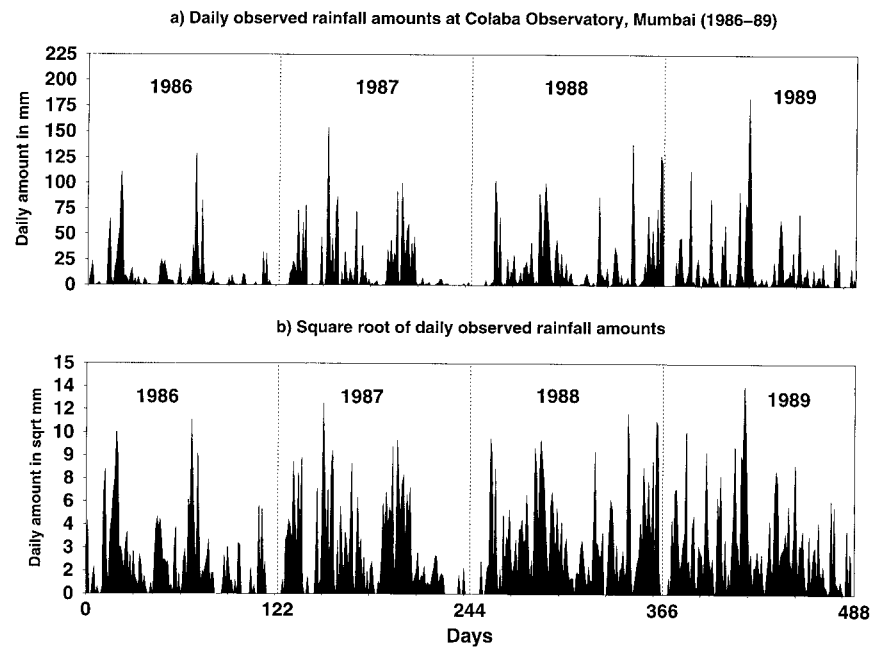b) Square root of daily observed rainfall amounts

FIG. 6. (a) Daily rainfall totals (in mm) measured at Colaba observatory (Mumbai) for June–September 1986–89, and (b) square root of the daily rainfall totals for the same period. Note the emergence of more low-amplitude structure in (b) compared to (a).

limiting case when $\nu \to 0^+$. Transformation strengths can be determined using many different methods (Box and Cox 1964). Aleksic and Jovanovic (1983) investigated the strength of the power law needed to minimize the sum of the squares of g_1 and g_2 for various gamma distributions and found that $\nu = 0.15$ was optimal for gamma distributions with a shape parameter of 0.5. The Weibull distribution suggests a strength of $\nu = \alpha/3.6$ to transform the data to be closest to Gaussian. (Stephenson et al. 1998b). Both these strength estimates are close to the fourth-root transformation $\nu = 0.25$ used by Simpson (1972) to test the significance of cloud seeding experiments. Power law transformations can also have the desirable effect of making the variance less dependent on the mean value, referred to as *variance stabilization* (Kendall et al. 1983). When the mean rainfall is small, the variance is also constrained to be small since rainfall amounts can never be negative. This can lead to nonstationary variations in the variance of time series (*heteroskedasticity*), which are not accounted for by simple autoregressive processes (e.g., persistence forecasts), or in power spectra analyses. The square root transformation is the optimal variance stabilizing transformation for a Poisson process and is, therefore, likely to be beneficial in stabilizing the variance of sporadic rainfall time series.

Our main aim is to reduce the relative contribution from extreme events, and so for the sake of simplicity we will focus on the intermediate strength $\nu = 0.5$ square root transformation that has previously been been successfully used to obtain reliable estimates of power spectra of intermittent non-Gaussian signals (Blackmon and Tukey 1958; Bloomfield 1975). Figure 6 shows the impact of this transformation on daily rainfall recorded at Mumbai for June–September 1986–89, where it can be noted that the transformed totals are less dominated by extreme events. The more Gaussian behavior is confirmed by smaller skewnesses around 1.2 after the transformation (Fig. 7) compared to typical skewnesses exceeding 3 of the raw data (Fig. 3d). The fourth-root transformation has also been tested on the observed In-
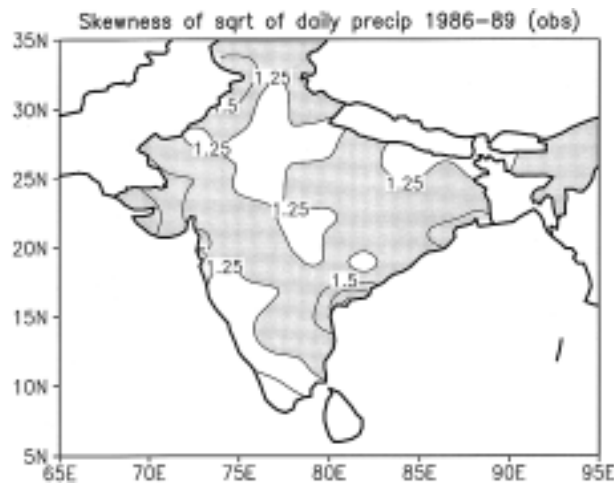

FIG. 7. Moment measure of skewness ($\sqrt{b_1}$) for the square root of daily wet-day rainfall amounts observed during June–September 1986–89. Note the marked reduction in skewness compared to that of the untransformed values shown in Fig. 3d.

dian rainfall totals and leads to even smaller values of skewness (about 0.5), yet at the expense of giving negative values of kurtosis (about -0.5) as to be expected from arguments based on the Weibull distribution (Stephenson et al. 1998b). We have therefore chosen to use the square root transformation in the rest of this study.

6. Generalizing the all-India rainfall index

Despite daily monsoon rainfall amounts differing enormously from one station to another, it is useful to summarize the general monsoon behavior by averaging over all the available station data to make an ''all-India rainfall'' index. Such an index has been calculated from 1871 to 1993 using the area-weighted average of 306 district stations covering the nonmountainous regions of India (Mooley and Parthasarathy 1984; Parthasarathy et al. 1994) and is defined as

$$\langle X \rangle = \frac{\sum_{i=1}^{n} A_i X_i}{\sum_{i=1}^{n} A_i}, \qquad (2)$$

where n is the daily varying number of stations having data, X_i is the rainfall daily total at station i, and the weight A_i is the area associated with station i. Area weighting aims to reduce the potential biases caused by inhomogeneities in the rain gauge network, yet in practice the area-weighted average differs only slightly from the simple arithmetic mean having equal weights. The all-India rainfall index is physically relevant for understanding the hydrological budget yet is not guaranteed to always be well sampled or always give reliable and meaningful statistical estimates. When one extreme local event can severely bias the overall mean value, as is the case for the all-India rainfall index, the representativeness of the mean can become highly debatable. In such cases, it can be advantageous to modify the physically motivated measure to be more statistically robust. One way of doing this is by using power law transformations to generalize the definition of mean to

$$\langle X \rangle_\nu = \langle X^\nu \rangle^{1/\nu} = \left(\frac{\sum_{i=1}^{n} A_i X_i^\nu}{\sum_{i=1}^{n} A_i} \right)^{1/\nu}, \qquad (3)$$

where ν is a strength parameter between zero and one. The familiar arithmetic mean is obtained for $\nu = 1$, and the geometric mean is obtained in the limit as $\nu \to 0^+$. Generalized means all summarize the average value but differ in the relative importance they give to large compared to small values. When ν is small, large values of X_i contribute less to the overall mean. A rational compromise between the arithmetic and the geometric means is obtained by taking $\nu = 0.5$, which defines the *square mean root* (smr). The smr is equivalent to taking the square root of the data before performing the arithmetic mean and then performing the inverse transform (squaring) on the resulting mean in order to regain the correct dimensional units (e.g., mm day^{-1}).

We have calculated three different all-India rainfall indices by gridding the irregular station data onto a regular 1° grid (using the algorithm described in section 2a) and then calculating the arithmetic mean, the square mean root, and the median of the rainfall amounts of the grid points lying within India's national borders. Both wet and dry days were included in the spatial averages. Figure 8 shows the daily time series of the three different indices for June–September of 1987 and 1988. The smr value is always smaller than the arithmetic mean value as to be expected from Jensen's inequality $\langle X \rangle_{\nu_1} \geq \langle X \rangle_{\nu_2}$ for $\nu_1 \geq \nu_2$ (Rao 1965). The smr is much closer to the median value than to the arithmetic mean and is less biased by extreme events such as those at the end of September 1988 (day 110), for example. The onset of the monsoon rainfall in June is more gradual in the smr and the median compared to that in the arithmetic mean. During the monsoon onset, small fractions of India are covered by often intense rainfall amounts that dominate the arithmetic mean but do not represent the overall relatively dry situation over India. The median and smr measures are more robust measures of the mean rainfall over the whole of India than is the arithmetic mean, which can be biased by strong local events. The intraseasonal oscillations in August 1987 (days 60–90) are also more clearly revealed in the smr and median measures than in the arithmetic mean. The smr offers some of the robustness advantages of the median yet with much less computational effort. The square root reduces the relative contribution of the local rapid extreme events and so might, therefore, be expected to lead to more persistence in the all-Indian rainfall. Figure 9 shows the autocorrelation functions for the arithmetic mean, median, and smr daily time series, estimated over the summers of 1986–89. The median and smr autocorrelation functions are similar and both have similar larger autocorrelations than does the arithmetic mean over time lags from 5 to 15 days. In other words, the signal is more persistent in the smr and median time series than in the arithmetic mean time series.

7. Ensemble forecasts of Indian rainfall

To test whether the square root transformation can improve general circulation model forecasts of Indian rainfall, we have performed an ensemble of nine forecasts for the month of June for each of the years 1986–89. June was chosen due to the marked differences noted between the arithmetic mean and the smr AIR indices for this month (Fig. 8). The ARPEGE model at T42 truncation was used to produce the forecasts and was briefly described in section 2b. A time-lagged approach was used to generate the ensemble members, with the atmosphere and land inital conditions set equal to the
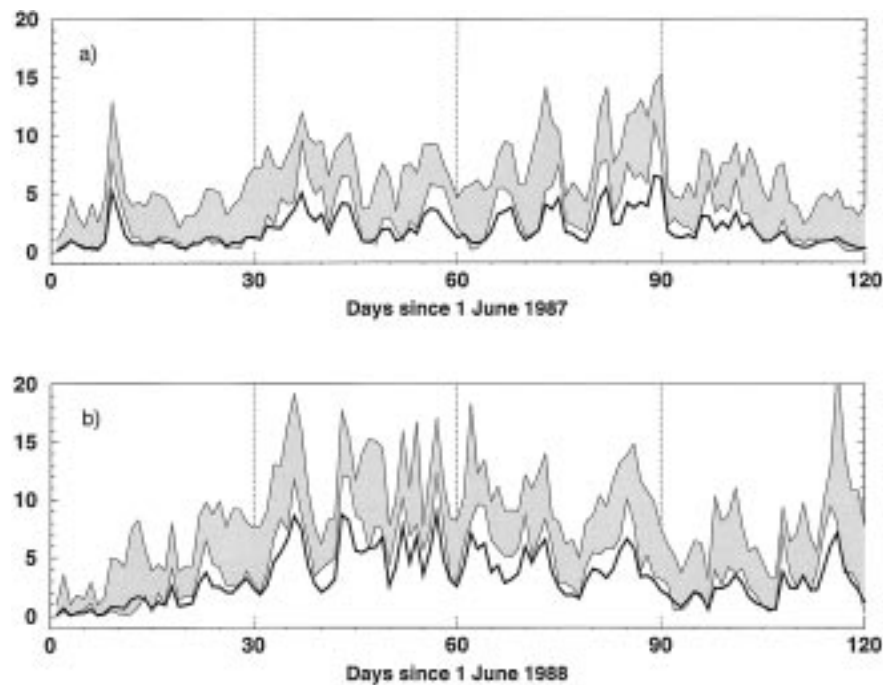
FIG. 8. All-India rainfall daily time series for the summers of (a) 1987 and (b) 1988. All-India rainfall indices have been estimated by taking the arithmetic mean (upper curve), the square mean root (middle curve), and the median (dark line) of the daily rainfall amounts at grid points over India. The gridded data on a 1° grid was obtained from the station data by the interpolation procedure described in section 2a.

ECWMF reanalyses at 1200 UTC on the 9 days preceding 1 June for each year (23–31 May). Sea surface temperatures were linearly interpolated for each day of the forecast between the monthly mean observed values.
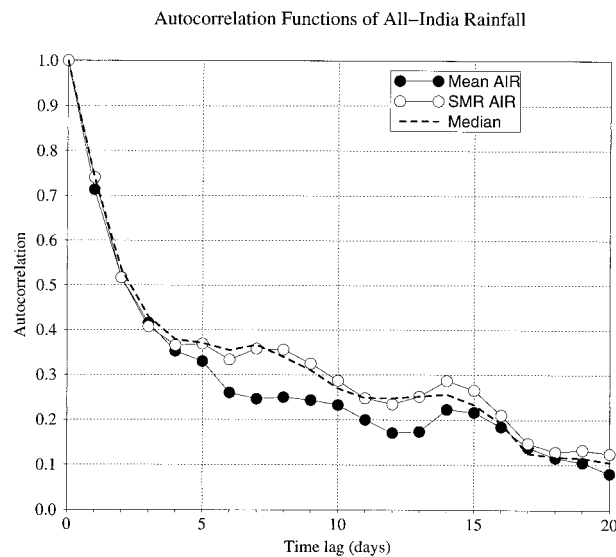

FIG. 9. Autocorrelation functions for the arithmetic mean, the smr, and the median daily AIR indices. An increase in autocorrelation can be discerned for lags of 5–15 days in both the smr and median estimates.

Daily Indian rainfall indices were calculated for each ensemble member extended-range forecast by taking both arithmetic mean and the smr of the 80 gridpoint values in the region covering India and its adjacent seas (5°–25°N, 70°–95°E). Monthly means were then obtained by taking the arithmetic mean of the indices over the 31 days in each June.

Arithmetic mean Indian rainfall forecasts for June 1986–89 are given in Table 2 and shown in Fig. 10a.

TABLE 2. June mean forecasts of arithmetic mean daily rainfall amounts (in mm day^{-1}) calculated over the 80 gridpoint values in the Indian region (5°–25°N, 70°–95°E). Mean is the arithmetic mean (in mm day^{-1}), and var is the interensemble variance of the ensemble values for each individual year [in (mm day^{-1})^2]. Obs is a verification value calculated similarly using ECMWF reanalyses of observations (in mm day^{-1}).

Forecast	1986	1987	1988	1989
23 May	7.63	5.99	8.21	5.27
24 May	8.18	6.91	7.21	6.05
25 May	9.00	7.04	7.02	6.30
26 May	7.66	7.22	7.98	6.56
27 May	9.17	6.18	7.09	8.16
28 May	7.60	6.16	9.66	8.98
29 May	7.57	8.23	6.75	7.29
30 May	7.74	8.06	6.25	7.44
31 May	8.13	9.10	8.11	6.86
Mean	8.08	7.21	7.59	6.99
Var	0.34	1.01	0.92	1.13
Obs	7.99	6.29	7.09	8.71

a) Arithmetic mean Indian rainfall in June

b) Square Mean Root Indian rainfall in June

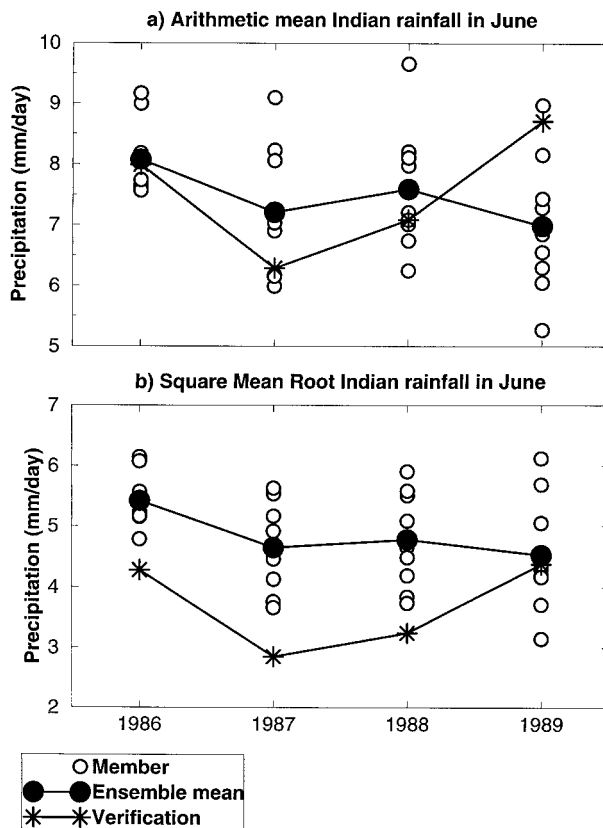O Member
● Ensemble mean
✳ Verification

FIG. 10. Ensemble forecasts of monthly mean rainfall averages for June 1986–89. (a) The monthly mean of daily arithmetic averages made over the 80 gridpoint values in the region (5°–25°N, 70°–95°E). (b) The monthly mean of the daily square mean root averages of the same 80 gridpoint values. Also depicted are the ECMWF verification reanalyses of observed data calculated in a similar fashion (*) and the arithmetic means of the nine ensemble values for each year (filled circles).

TABLE 3. June mean forecasts of square mean root daily rainfall amounts (in mm day^{-1}) calculated over the 80 gridpoint values in the Indian region (5°–25°N, 70°–95°E). Mean is the mean (in mm day^{-1}), and var is the interensemble variance of the ensemble values for each individual year [in (mm day^{-1})^2]. Obs is a verification value calculated similarly using ECMWF reanalyses (in mm day^{-1}).

Forecast	1986	1987	1988	1989
23 May	4.79	4.13	5.51	3.15
24 May	5.42	4.58	4.66	3.71
25 May	6.15	4.92	4.19	4.19
26 May	5.23	4.46	5.09	4.31
27 May	6.08	3.76	4.49	5.69
28 May	5.16	3.66	5.91	6.13
29 May	5.24	5.54	3.84	5.06
30 May	5.17	5.17	3.74	4.38
31 May	5.57	5.63	5.59	4.17
Mean	5.42	4.65	4.78	4.53
Var	0.18	0.47	0.55	0.78
Obs	4.28	2.85	3.24	4.38

A verification analysis has also been included by calculating an index in an identical fashion using the daily ECMWF rainfall reanalyses interpolated spectrally onto the T42 model grid. There is a wide spread between the ensemble forecasts in each of the years and this can be considered to be a form of unpredictable sampling noise caused by individual weather events. Outlier forecasts are also clearly apparent, such as the extremely wet forecast in 1988, which biased the mean toward a higher value. Potential predictability can be quantified by calculating the signal-to-noise F ratio of the interannual variance of the means of the ensemble forecasts to the mean of the interensemble variances for each year (Madden 1976; Leith 1978). The interannual variance of the ensemble means (0.41 mm day^{-1})^2, is considerably less than the 1986–89 mean of the interensemble variances (0.92 mm day^{-1})^2, resulting in a small F ratio of 0.20. There is much less variance in the predictable signal than there is in the sampling noise. The F ratios are smaller than those of Shea et al. (1995) because forecasts are made here only for one month rather than for a whole

season, and it is also quite likely that the model underestimates the interannual signal in the Indian region.

Table 3 and Figure 10b present ensemble forecasts of the square mean root of the gridded Indian rainfall amounts. There are fewer outliers in the smr forecasts than were present in the forecasts of the arithmetic mean. The interannual variance of the means of the forecasts is (0.35 mm day^{-1})^2 compared to the mean interensemble variance of (0.70 mm day^{-1})^2, and therefore the F ratio of the smr forecasts is 0.25. While this ratio is still small indicating not much potential predictability, it is encouraging to note that the F ratio is 25% larger than that for the arithmetic means. This increase results from having both more persistence in the daily time series (as shown in section 6), and also having more reliable smaller estimates of the interensemble variance due to having less outlying forecasts. More reliable estimates of the interensemble variance could improve the noted poor ability of ensemble forecasts to predict the spread in forecast rainfall amounts (Hammill and Colucci 1998).

Despite the potential predictability being small, and the period 1986–89 being short, it is interesting to ask whether or not the skill of interannual forecasts of smr Indian rainfall was higher than that of forecasting the arithmetic mean Indian rainfall. If one judges the forecast skill by the correlation between the verification analyses and the mean of the ensemble forecasts, one finds some skill for the smr forecasts ($r = 0.35$) rather than almost no skill for the forecasts of the arithmetic mean ($r = -0.01$). Most of this gain in skill comes from an improvement in the poor mean forecast in June 1989 and suggests that the square root transformation of daily rainfall amounts can improve the skill of rainfall forecasts. A possible explanation for this improvement is suggested by the study of D'Amato and Lebel (1998), which demonstrated that interannual variability of Sahel rainfall is linked to the number rather than the magnitude of the rainfall events. The square root transformation

can also be beneficial for forecasts over either smaller regions or shorter periods, such as statistical predictions of regional rainfall (Feddersen et al. 1999). The square root transformation gives less weight to anomalies in large rainfall amounts and can be considered as equivalent to using a non-Euclidean positively curved measure of forecast skill (Stephenson 1997).

8. Concluding remarks

Daily rainfall amounts are often extreme events many standard deviations above the expected mean value, and such outliers can cause large sampling errors in estimated rainfall statistics. The spatial distribution of Indian rainfall extreme events has been analyzed in this study using statistics such as the coefficient of variation and skewness. The most extreme variability occurs over the coastal regions of eastern and northwestern India. Similar values yet different spatial patterns are obtained for daily rainfall data generated by the ARPEGE atmospheric model. We have also demonstrated that while gamma and Weibull distributions provide reasonable fits, the lognormal distribution is too skewed to provide a good fit to the low-order moments of Indian wet-day daily rainfall totals.

Evidence has been presented showing how individual daily rainfall events can have a large impact even on seasonal mean rainfall amounts. Higher-order statistics such as variances and covariances are even less robust than means in the presence of extreme events, and therefore extreme care should be exercised when interpreting such quantities based on rainfall amounts. For example, eigenvectors of the estimated rainfall-covariance have been used in previous studies to isolate dominant modes of monsoon intraseasonal variability, yet such quantities explain only a small fraction of the total variance and are poorly estimated due to the the presence of outliers (Ferranti et al. 1997 and references therein). Estimates of power and singular spectra are also prone to sampling errors, which can give rise to spurious peaks especially on intraseasonal and shorter timescales. Fortunately, a simple method exists for reducing the relative contribution from extreme events. By nonlinearly transforming the daily rainfall amounts, the resulting time series become closer to Gaussian and contain less troublesome extreme events. The square root transformation is easy to apply and appears to work well for Indian rainfall. It has the advantage over the logarithm transformation that it is not singular for zero rainfall amounts. Such a transformation is easily applied to daily data before performing the desired statistical analyses and helps make the results more robust. The advantages of using more robust statistics are reviewed in Lanzante (1996).

When forecasting climatic quantities, it is important to be aware of how robust the quantities are in the presence of sampling fluctuations caused by individual weather events. For example, area averages over large regions are generally more robust than averages over small regions and, therefore, can generally be forecast with more skill. Not all quantities of human interest are sufficiently robust to be predicted accurately and one should therefore focus attention on robust quantities if one wishes to avoid poor predictions. Because of the immense complexity of the climate system, it will not be possible to produce accurate forecasts for all quantities of interest, yet it may sometimes be possible to transform variables to obtain quantities that can be reliably predicted. The arguments presented in this article are general and are, therefore, applicable to other severely skewed quantities wherever they occur.

REFERENCES

Aleksic, N., and M. Jovanovic, 1983: On power transformations of gamma distributed variables. Preprints, *Eighth Conf. on Probability and Statistics in Atmospheric Sciences,* Hot Springs, AR, Amer. Meteor. Soc., 1–2.

Barger, G. L., and H. C. S. Thom, 1949: Evaluation of drought hazard. *Agron. J.,* **41,** 519–526.

Blackman, R. B., and J. W. Tukey, 1958: *The Measurement of Power Spectra.* Dover, 190 pp.

Bloomfield, P., 1975: *Fourier Analysis of Time Series: An Introduction.* Wiley-Interscience, 258 pp.

Box, G. E. P., and D. R. Cox, 1964: An analysis of transformations. *J. Roy. Stat. Soc.,* **B26,** 211–243.

Charney, J. G., and J. Shukla, 1981: Predictability of monsoons. *Monsoon Dynamics,* J. Lighthill, Ed., Cambridge University Press, 99–110.

Cressman, G. P., 1959: An operational objective analysis scheme. *Mon. Wea. Rev.,* **87,** 329–340.

D'Amato, N., and T. Lebel, 1998: On the characteristics of the rainfall events in the Sahel with a view to the analysis of climatic variability. *Int. J. Climatol.,* **18,** 955–974.

Dolby, J. L., 1963: A quick method of choosing a transformation. *Technometrics,* **5,** 317–334.

Essenwanger, O. M., 1986: Frequency distribution. *General Climatology: Elements of Statistical Analysis,* H. E. Landsberg, Ed., World Survey of Climatology, Vol. 1B, Elsevier, 28–101.

Feddersen, H., A. Navarra, and M. N. Ward, 1999: Reduction of model systematic error by statistical correction for dynamical seasonal predictions. *J. Climate,* **12,** 1974–1989.

Ferranti, L., J. M. Slingo, T. N. Palmer, and B. J. Hoskins, 1997: Relations between interannual and intraseasonal monsoon variability as diagnosed from AMIP integrations. *Quart. J. Roy. Meteor. Soc.,* **123,** 1323–1357.

Haldane, J. B. S., 1938: The approximate normalization of a class of frequency distributions. *Biometrika,* **29,** 392–407.

Hamill, T. M., and S. J. Colucci, 1998: Evaluation of Eta-RSM ensemble probabilistic precipitation forecasts. *Mon. Wea. Rev.,* **126,** 711–724.

Ison, N. T., A. M. Feyerherm, and L. D. Bark, 1971: Wet period precipitation and the gamma distribution. *J. Appl. Meteor.,* **10,** 658–665.

Katz, R. W., 1983: Statistical procedures for making inferences about precipitation changes simulated by an atmospheric general circulation model. *J. Atmos. Sci.,* **40,** 2193–2201.

Kedem, B., L. S. Chiu, and G. R. North, 1990: Estimation of mean rain rate: Application to satellite observations. *J. Geophys. Res.,* **95,** 1965–1972.

——, H. Pavlopoulos, X. Guan, and D. A. Short, 1994: A probability distribution model for rain rate. *J. Appl. Meteor.,* **33,** 1486–1493.

Kendall, M., A. Stuart, and J. K. Ord, 1983: *The Advanced Theory of Statistics.* 4th ed. Macmillan, 780 pp.

Keshavamurty, R. N., 1973: Power spectra of large-scale disturbances of the Indian southwest monsoon. *Ind. J. Meteor. Geophys.,* **24,** 117–124.

Krishnamurti, T. N., M. Kanamitsu, R. V. Godbole, C. B. Chang, F. Carr, and J. H. Chow, 1975: Study of a monsoon depression. (I) Synoptic structure. *J. Meteor. Soc. Japan,* **53,** 227–240.

Lal, M., L. Bengtsson, U. Cubasch, M. Esch, and U. Schlese, 1995: Synoptic scale disturbances of Indian summer monsoon as simulated in a high resolution climate model. *Climate Res.,* **5,** 243–258.

Lanzante, J. R., 1996: Resistant, robust and non-parametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data. *Int. J. Climatol.,* **16,** 1197–1226.

Leith, C. E., 1978: Predictability of climate. *Nature,* **276,** 352–355.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.,* **20,** 130–141.

Madden, R. A., 1976: Estimates of the natural variability of time-averaged sea-level pressure. *Mon. Wea. Rev.,* **104,** 942–952.

Mooley, D. A., 1973a: Gamma distribution probability model for Asian summer monsoon monthly rainfall. *Mon. Wea. Rev.,* **101,** 160–176.

——, 1973b: An estimate of the distribution and stability period of the parameters of the gamma probability model applied to monthly rainfall over Southeast Asia during the summer monsoon. *Mon. Wea. Rev.,* **101,** 884–890.

——, and H. L. Crutcher, 1968: An application of gamma distribution function to Indian rainfall. ESSA Tech. Rep. 5, Environmental Data Service, Silver Springs, MD, 47 pp. [Available from U. S. Department of Commerce, Springfield, VA 22161.]

——, and G. Appa Rao, 1970: Statistical distribution of pentad rainfall over India during monsoon season. *Indian J. Meteor. Geophys.,* **21,** 219–230.

——, and ——, 1971: Distribution function for seasonal and annual rainfall over India. *Mon. Wea. Rev.,* **99,** 796–799.

——, and B. Parthasarathy, 1984: Fluctuations in all-India summer monsoon rainfall during 1871–1978. *Climate Change,* **6,** 287–301.

Osborn, T. J., and M. Hulme, 1997: Development of a relationship between station and grid-box rainday frequencies for climate model evaluation. *J. Climate,* **10,** 1885–1908.

Öztürk, A., 1981: On the study of a probability distribution for precipitation totals. *J. Appl. Meteor.,* **20,** 1499–1505.

Palmer, T. N., 1994: Chaos and predictability in forecasting the monsoons. *Proc. Indian Natl. Sci. Acad.,* **60A,** 57–66.

Pant, G. B., K. Rupa Kumar, B. Parthasarathy, and H. P. Borgaonkar, 1988: Long term variability of the Indian summer monsoon and related parameters. *Adv. Atmos. Sci.,* **5,** 469–481.

Parthasarathy, B., A. A. Munot, and D. R. Kothawale, 1994: All-India monthly and seasonal rainfall series: 1871–1993. *Theor. Appl. Climatol.,* **49,** 217–224.

Pearson, E. S., and H. O. Hartley, 1972: *Biometrika Tables for Statisticians.* Vol. 2. Cambridge University Press, 65 pp.

Pramanik, S. K., and P. Jagannathan, 1953: Climatic changes in India. (I)—Rainfall. *Indian J. Meteor. Geophys.,* **4,** 291–309.

Ramage, C. S., 1971: *Monsoon Meteorology.* International Geophysics Series, Vol. 15, Academic Press, 296 pp.

Rao, C. R., 1965: *Linear Statistical Inference and its Applications.* Wiley Publications, 625 pp.

Rodwell, M. J., 1997: Breaks in the Asian monsoon: The influence of Southern Hemisphere weather systems. *J. Atmos. Sci.,* **54,** 2597–2611.

Ropelewski, C. F., and J. B. Jalickee, 1983: Estimating the significance of seasonal prediction amounts using approximations of the inverse gamma function over an extended range. Preprints, *Eighth Conf. on Probability and Statistics in Atmospheric Sciences,* Hot Springs, AR, Amer. Meteor. Soc., 125–129.

——, J. E. Janowiak, and M. S. Halpert, 1985: The analysis and display of real-time surface climate data. *Mon. Wea. Rev.,* **113,** 1101–1106.

Sankaranarayanan, D., 1933: On the nature of frequency distribution of precipitation in India during monsoon months June to September. *Indian Meteor. Dept. Sci. Notes,* **5,** 97–107.

Shea, D. J., and N. A. Sontakke, 1995: The annual cycle of precipitation over the Indian subcontinent: Daily, monthly, and seasonal statistics. NCAR Tech. Note NCAR/TN-401 + STR, 168 pp. [Available from NCAR, P.O. Box 3000, Boulder, CO 80307.]

——, ——, R. A. Madden, and R. W. Katz, 1995: The potential for long-range prediction of precipitation over India for the Southwest monsoon season: An analysis of variance approach. *Proc. Sixth Int. Meeting on Statistical Climatology,* Galway, Ireland, Steering Committee for the International Meetings on Statistical Climatology, 475–478.

Shenton, L. R., and K. O. Bowman, 1973: Note on the sample size to achieve normality for estimators for the gamma distribution. *Mon. Wea. Rev.,* **101,** 891–892.

Sikka, D. R., 1977: Some aspects of the life history, structure and movement of monsoon depressions. *Pure Appl. Geophys.,* **115,** 1501–1529.

Simpson, J., 1972: Use of the gamma distribution in single-cloud rainfall analysis. *Mon. Wea. Rev.,* **100,** 309–312.

Skelly, W. C., and A. Henderson-Sellers, 1996: Grid box or grid point: What type of data do GCMs deliver to climate impacts researchers? *Int. J. Climatol.,* **16,** 1079–1086.

Soman, M. K., and K. Krishna Kumar, 1990: Some aspects of daily rainfall distributions over India during southwest monsoon season. *Int. J. Climatol.,* **10,** 299–311.

Stephenson, D. B., 1997: Correlation of spatial climate/weather maps and the advantages of using the Mahalanobis metric in predictions. *Tellus,* **49A,** 513–527.

——, F. Chauvin, and J. F. Royer, 1998a: Simulation of the Asian summer monsoon and its dependence on model horizontal resolution. *J. Meteor. Soc. Japan,* **76,** 237–265.

——, K. Rupa Kumar, F. J. Doblas-Reyes, J. F. Royer, and F. Chauvin, 1998b: Extreme daily rainfall events and their impact on the potential predictability of the Indian monsoon. Note de Travail de Météo-France GMGEC 63, 41 pp. [Available from Dr. D. B. Stephenson, Laboratoire de Statistique et Probabilitiés, Université Paul Sabatier, 118, Route de Narbonne, F-31062 Toulouse Cedex, France.]

Thom, H. C. S., 1958: A note on the gamma distribution. *Mon. Wea. Rev.,* **86,** 117–122.

Tukey, J. W., 1955: On the comparative anatomy of transformations. *Ann. Math. Stat.,* **28,** 602–632.

Weibull, W., 1951: A statistical distribution function of wide applicability. *J. Appl. Mech.,* **18,** 293–297.

Wilheit, T. T., A. T. C. Chang, and L. Chiu, 1991: Retrieval of monthly rainfall indices from microwave radiometeric measurements using probability distribution functions. *J. Atmos. Oceanic Technol.,* **8,** 118–136.

Wilks, D. S., 1989: Rainfall intensity, the Weibull distribution, and estimation of daily surface runoff. *J. Appl. Meteor.,* **28,** 52–58.

Wong, R. K. W., 1977: Weibull distribution, iterative likelihood techniques and hydrometeorological data. *J. Appl. Meteor.,* **16,** 1360–1364.