Assessment of temporal change in the tails of probability distribution of daily precipitation over India due to climatic shift in the 1970s

Neha Gupta and Sagar Rohidas Chavan

ABSTRACT

Daily precipitation extremes are crucial in the hydrological design of major water control structures and are expected to show a changing tendency over time due to climate change. The magnitude and frequency of extreme precipitation can be assessed by studying the upper tail behavior of probability distributions of daily precipitation. Depending on the tail behavior, the distributions can be classified into two categories: heavy-tailed and light-tailed distributions. Heavier tails indicate more frequent occurrences of extreme precipitation events. In this paper, we have analyzed the temporal change in the tail behavior of daily precipitation over India from pre- to post-1970 time periods as per the global climatic shift. A modified Probability Ratio Mean Square Error norm is used to identify the best-fit distribution to the tails of daily precipitation among four theoretical distributions (e.g., Pareto-type II, Lognormal, Weibull, and Gamma distributions). The results indicate that the Lognormal distribution, which is a heavy-tailed distribution, fits the tails of daily precipitation for the majority of the grids. It is inferred from the study that there is an increase in the heaviness of tails of daily precipitation data over India from pre- to post-1970 time periods.

Key words | daily precipitation extremes, global climatic shift, heavy-tailed distribution, light-tailed distribution

Neha Gupta (corresponding author)
Sagar Rohidas Chavan
Department of Civil Engineering,
Indian Institute of Technology Ropar,
Rupnagar, Punjab 140001,
India
E-mail: *2017cez0006@iitrpr.ac.in*

HIGHLIGHTS

- This paper attempts to assess the changes in the tail behavior of daily precipitation over India due to climatic shift in the 1970s.
- Maps show best-suited probability distributions for pre-/post-1970s and highlight the changes.
- Results highlight the use of heavy-tailed probability distributions for analyzing the extreme precipitation.
- A significant increase in heaviness of tails is noticed from pre- to post-1970 time periods.

INTRODUCTION

Among many of the well-known consequences of climate change, the intensification of hydrological variables such as precipitation and temperature is significant. Climate change leads to an increase in the intensity and frequency of rainfall events globally and locally (Groisman *et al.* 1999; Jacob & Hagemann 2007; Giorgi *et al.* 2011). The Intergovernmental Panel on Climate Change (IPCC) has put serious efforts into summarizing the impacts of extreme precipitation on floods in the Special Report on Extremes

doi: 10.2166/wcc.2021.008

(SREX) (IPCC 2013). This shows the grave concern of the hydro-meteorologists worldwide about the hydro-climatic extremes and their consequences for mankind. The reliable estimation of daily extreme precipitation events is of utmost importance to ensure the structural safety of infrastructure projects and prevent loss of human life (Koutsoyiannis 2008). These rare, abnormal, and catastrophic events usually lie in the tail part of the probability distribution of daily precipitation data. The tail is defined as the upper part of the complementary cumulative distribution function beyond a particular threshold (e.g., Klüppelberg 1988). The conventional distribution fitting methods are unable to fit the tail of daily precipitation data adequately, which results in the exemption of extreme precipitation events (termed as improbable events or outliers), thereby underestimating their probability of occurrence (Koutsoyiannis 2004a, 2004b; Li *et al.* 2012; Papalexiou *et al.* 2013; Chen & Brissette 2014; Beskow *et al.* 2015; Papalexiou & Koutsoyiannis 2016; De Michele & Avanzi 2018; Mlyński *et al.* 2019). As per the classical extreme value theory, the block maxima (BM) extracted from a time-series resemble one of the three limiting distributions, namely, (i) Gumbel distribution (i.e., Extreme Value Type I distribution); (ii) Fréchet distribution (i.e., Extreme Value Type II distribution); and (iii) reversed Weibull (i.e., Extreme Value Type III distribution) (Fisher & Tippett 1928; Gnedenko 1943; Jenkinson 1955; Coles 2001; Langousis *et al.* 2016). In the case of annual maximum daily precipitation (i.e., BM per year), Gumbel and Fréchet distributions were found to be appropriate to model the behavior of extremes as both possess unbounded upper tail behavior (Koutsoyiannis 2004a; Papalexiou & Koutsoyiannis 2013; De Michele 2019). Generally, the reversed Weibull distribution is not considered for analyzing annual maximum daily precipitation as the distribution is bounded on the upper tail (Kotz & Nadarajah 2000; Papalexiou & Koutsoyiannis 2013; Chavan & Srinivas 2021). The inferences from the BM approach are dependent on the selection of block size (i.e., either annual or seasonal maxima, etc.). The selection of the annual maxima (AM) from daily precipitation records at a location may distort the tail behavior of their probability distribution as it might miss a few of the largest daily precipitation events from a particular year. As the BM approach discards a large portion of information from available data, the estimated distribution parameter exhibits significant variability and becomes sensitive to outliers (Coles *et al.* 2003; Koutsoyiannis 2004b; Deidda 2010; Langousis *et al.* 2016). Another way of modeling extreme precipitation is based on the peak-over-threshold (POT) approach. In the POT approach, a sample is extracted from daily precipitation series by selecting all observations above an arbitrary threshold u (Chow 1964). As the threshold increases, such samples tend to follow generalized Pareto distribution (GPD) (e.g., Balkema & de Haan 1974; Pickands 1975; Madsen *et al.* 1997). Many studies from the past revealed that the findings based on the POT approach are generally more efficient than the BM approach (Cunnane 1973; Madsen *et al.* 1997; Caires 2009; Villarini *et al.* 2011; Moccia *et al.* 2021). Despite its advantages, the use of POT is less prevalent than BM due to (i) the presence of serial dependence in identified peaks and (ii) ambiguity in the selection of an optimum threshold for the identification of peaks (e.g., Beguería 2005; Mailhot *et al.* 2013; Serinaldi & Kilsby 2014; Kiran & Srinivas 2021). To avoid the loss of information about the extreme precipitation as in the case of the BM approach and the selection of arbitrary threshold as in the case of the POT approach, an annual exceedance series (AES) is used for demarcating the tail of daily precipitation data in this study. An AES approach has the advantage of better representing the exact tail of the parent distribution (e.g., Chow 1964; Ben-Zvi 2009; Gupta 2011; Papalexiou *et al.* 2013). After finalizing the tail of daily precipitation data, the distribution fitting can be accomplished using the probability ratio mean square error (PRMSE) norm proposed by Papalexiou *et al.* (2013). PRMSE norm-based tail-fitting of the probability distributions yields unbiased estimates of parameters of distributions and also compares the fits of various probability distributions (e.g., Papalexiou *et al.* 2013, 2018; Moccia *et al.* 2021). In a recent study by Moccia *et al.* (2021), the equivalence between PRMSE and a conventional fitting method, i.e., Kolmogorov–Smirnov (KS) test, in identifying the best-fitting distribution was verified, and the PRMSE was found to have some additional benefits over the KS test.

Classification of probability distributions based on their tail behavior

Various probability distributions can be classified based on their tail behavior into two categories: heavy-tailed distributions (e.g., Pareto, Lognormal, Weibull, Lévy, etc.) and light-tailed distributions (e.g., Gaussian, Exponential, etc.). Heavy-tailed distributions are also referred to as 'fat-tailed', 'thick-tailed,' or 'long-tailed' in various literatures (El Adlouni *et al.* 2008; Foss *et al.* 2013; Papalexiou *et al.* 2013). They have upper tails decaying as a power law (i.e., tails tend to approach zero more gently than an exponential tail). A random variable X is said to have a heavy-tail when its moment-generating function becomes infinite on R (the set of real number) given in Equation (1) (e.g., Panorska *et al.* 2007; Foss *et al.* 2013; Panahi 2016; Wang *et al.* 2018).

$$\int_R e^{-\lambda x} F(x) dx = \infty \text{ for all } \lambda > 0 \tag{1}$$

Heavy-tailed distributions belong to a class of subexponential distributions. The class of subexponential distribution was initially introduced by Chistyakov (1964), and for any distribution function F to be subexponential, one of the following conditions must hold.

$$\text{(a)} \quad \lim_{x \to \infty} \frac{\bar{F}^{n*}(x)}{\bar{F}(x)} = n \text{ for some (all) } n \geq 2 \tag{2}$$

$$\text{(b)} \lim_{x \to \infty} \frac{P(X_1 + \ldots + X_n > x)}{P(\max(X_1, \ldots, X_n) > x)} = 1 \text{ for some (all) } n \geq 2 \tag{3}$$

where $\bar{F}(x)$ denotes the exceedance probability; $\bar{F}^{n*} = 1 - F^{n*}(x) = P(X_1 + X_2 + \ldots\ldots + X_n > x)$ denotes the tail of n-fold convolution of F (Embrechts & Goldie 1982; Embrechts *et al.* 1997; Goldie & klüppelberg 1998). Definition (a) shows the absence of any exponential moments, while condition (b) suggested by Teugel (1975) indicates the presence of enormous values (i.e., rare events) in the sample. Condition (b) is also known as the principle of 'a single big jump' (Foss *et al.* 2013; Hill 2019). Embrechts & Goldie (1980) showed that the equivalence of condition (a) and (b) power-law distributions like Lognormal distribution and regularly varying distributions such as Pareto-type II distribution are subsets of the subexponential distributions (Feller 1971; Bingham *et al.* 1987; El Adlouni *et al.* 2008; Foss *et al.* 2013; Voitalov *et al.* 2018). Several attempts have been made to group tails of distributions according to their limiting/asymptotic behavior (e.g., Goldie & Klüppelberg 1998; Ouarda *et al.* 1994, etc.). Werner & Upper (2004) classified the distributions in five nested classes from A to E such that A⊂ B⊂ C⊂ D⊂ E: Class A (stable distributions; distributions with Pareto tails having $\alpha < 1$), B (Pareto-type tail), C (regularly varying distributions), D (subexponential distributions), and the class E (the broadest class includes distributions with infinite exponential moments). El Adlouni *et al.* (2008) combined classifications mentioned above with five graphical criteria for tail discrimination and arranged different distributions from light- to heavy-tailed. Papalexiou *et al.* (2013, 2018) intuitively defined two broad classes of distributions based on the asymptotic tail behavior as (a) subexponential class (heavy-tailed class) and (b) the superexponential class (Nagaev & Tsitsiashvili 2006) or hyperexponential class (Vela & Rodríguez 2014) (light-tailed class). The practical implication of a heavy-tailed distribution such as a Pareto or a lognormal distribution is that the large values representing rare events are much more likely to occur than that of a light-tailed distribution like Gaussian or Exponential distribution.

Several studies like that of Mielke (1973) have shown that heavy-tailed distributions like Kappa distribution might be more suitable for modeling daily precipitation data. Panorska *et al.* (2007) verified that the daily extreme precipitation data come from Pareto distribution rather than an exponential distribution based on likelihood ratio tests for hypothesis validation. They found that the tails of daily precipitation closely resemble power law for most North American continent stations. Papalexiou *et al.* (2013) used the daily rainfall of 15,137 records and compared the performance of various probability distributions such as Pareto-type II, Weibull, Lognormal, and Gamma distributions in describing the upper tails and ranked the suitability of the distributions as per PRMSE norm. Overall findings from the study indicated that the daily rainfall extremes are better described by heavy-tailed distribution. Cavanaugh *et al.* (2015) used the statistical test of Kozubowski *et al.* (2009) and the methodology of Panorska *et al.* (2007) to show that the probability distribution of intense daily precipitation over 22,000 stations located globally exhibits heavy tails. The Pareto-type tail dominates over 65% of stations as compared with the exponential tail-type distributions. Nerantzaki & Papalexiou (2019) developed a faster algorithmic procedure for the mean excess function (MEF)-based graphical method to discriminate between exponential and subexponential tails for about 21,348 daily precipitation records all over the globe. They observed that nearly 75.8% of records showed the dominance of heavy-tail distributions. Graphical methods such as Log–log plots, MEF, Zipf plot, and Hill tail estimator are also useful in assessing the tail behavior of distributions (Hill 1975; Embrechts *et al.* 1997; Ghosh & Resnick 2010; Nieboer 2011; Cirillo 2013; Cooke *et al.* 2014). Recently, Moccia *et al.* (2021) showed that heavy-tailed distributions provide a better fit than the light-tailed distributions for about 80% of stations in two Italian regions (i.e., Lazio and Sicily). They perform the analysis for the sample selected through the AM and AES approaches.

Effect of global climatic shift on extreme precipitation

There have been several studies that analyzed the changes in extreme rainfall over India at the national and regional scale and have drawn different conclusions (Goswami *et al.* 2006; Rajeevan *et al.* 2008; Dash *et al.* 2009; Ghosh *et al.* 2012; Kulkarni *et al.* 2012; Mondal & Mujumdar 2015; Shastri *et al.* 2015; Ghosh *et al.* 2016; Singh *et al.* 2016; Roxy *et al.* 2017; Bisht *et al.* 2018a, 2018b). The changes in extreme precipitation can be attributed to the abrupt global change of the climatic system caused by a regime shift in the 1970s in various climatic factors like the Arctic Oscillation (AO), East Asian summer monsoon (EASM), East Asian winter monsoon (EAWM), El Niño–Southern Oscillation (ENSO), North Atlantic Oscillation (NAO), Aleutian low (AL), Pacific decadal oscillation (PDO), Western Pacific subtropical high (WPSH), and Indian summer monsoon rainfall (ISMR) (Biondi *et al.* 2001; Chowdary *et al.* 2006; Zhou *et al.* 2009; O'Kane *et al.* 2014; Chen *et al.* 2015; Sahana *et al.* 2015; Hsu 2016; Zuo *et al.* 2016; Weisheimer *et al.* 2017; Dai *et al.* 2018) or some local changes such as urbanization. The climate regime shift has adversely impacted the atmosphere, ecosystems, biological, and many hydro-climatic variables, such as temperature, air pressure, wind field, and rainfall, resulting in the frequent occurrence of extremes like heat, drought, heavy rainfall, and flood disasters (Graham 1994; Zhang *et al.* 1998; Wang 2001; Meehl *et al.* 2008; Jacques-coper & Garreaud 2015; Huang *et al.* 2017a, 2017b). Many researchers have demonstrated that the rainfall characteristics within and beyond the monsoon period exhibit spatio-temporal changes due to climate regime shifts globally and in India (Sabeerali *et al.* 2012; Sahana *et al.* 2015). Ajayamohan & Suryachandra (2008) have also shown an increased extreme rainfall event over central India after the 1976/1977 climate shift. Vittal *et al.* (2013) showed that rainfall extremes have changed in India after 1975 and established that urbanization, in terms of change in population density, is a possible cause of change. They used a comprehensive POT approach with 95 and 99 percentile thresholds, including multiple extreme events in a year. Dash & Maity (2019) found that the precipitation-based climate change indices exhibit increasing trends over India with more spatial extent post-1975. Recently, Sarkar & Maity (2020) observed an increment of 35% in probable maximum precipitation over India in the post-1970 (1971–2010) period when compared with the pre-1970 (1901–1970) period due to climate shift.

Study goals

This study proposes to assess the temporal changes in the daily extreme precipitation over India due to climatic shift in the 1970s based on the change in the tail behavior of the probability distributions of daily precipitation. We intend to perform the assessment by considering four theoretical distributions (e.g., Pareto-type II, Lognormal, Weibull, and Gamma distributions), which belong to different classes of distributions (i.e., subexponential or hyperexponential classes) following Papalexiou *et al.* (2013). The primary goals of this study are (a) to find the best-suited distribution based on the PRMSE norm that can describe the extreme daily precipitation in changing climate over India, (b) to provide a categorical classification of grids into two broad classes of distribution, i.e., subexponential class and the hyperexponential–exponential class considering the shift in the global climatic regime in the 1970s; (c) to investigate spatial and temporal changes in the behavior of tails of the probability distribution of daily precipitation between the two time periods, namely pre-1970 (1901–1970) and post-1970 (1971–2010). The assessment is performed both at the grid and regional scale (i.e., Meteorological Subdivisions). The study would be useful to the design engineers and hydro-meteorologists for reliable planning and management of various major water-energy infrastructures in India.

DATA AND METHOD

Study area

India, the largest South Asian country with a wide variety of climatic regions extending from low rainfall arid regions to the heavy rainfall receiving regions, is our study area. The climatic condition of the Indian mainland is influenced by various geographical and relief features like the Himalayas in the north, Thar Desert and Arabian Sea in the west, the Bay of Bengal in the east, Western Ghats in the southwest, and the Indian Ocean in the south. The study area covers a widespread range of variations in the rainfall extremes, which motivates us to examine spatial and temporal behavior of daily extreme precipitation in terms of magnitude and frequency of occurrence between the two time periods corresponding to the shift in the global climate regime in the 1970s, i.e., pre-1970 and post-1970. Furthermore, the temporal changes are investigated at a regional scale in 34 out of 36 homogeneous Meteorological Subdivisions (see Figure S1 in the Supplementary Material) in this analysis (Guhathakurta & Rajeevan 2008).

Rainfall data

In this study, an extensive database of daily gridded precipitation with a spatial resolution of 0.25° procured from the India Meteorological Department (IMD) is considered. The gridded rainfall data were prepared for 112 years (1901–2013) by Pai *et al.* (2014) using a varying network of 6,955 rain gauge stations. After performing a quality check, 4,789 grids each having a record length (N) of 110 years, i.e., from 1901 to 2010, were selected for analysis. No missing data were filled at any grids/stations. Records at each grid have been split into two parts, i.e., pre-1970 (1901–1970) and post-1970 (1971–2010), to capture the effect of the shift in the global climate regime. Despite being an unequal division of the data, the records at each grid for both pre- and post-1970 time periods have a sufficient number of non-zero daily precipitation values needed to estimate the tail behavior using the threshold-based approach for fitting probability distributions to daily precipitation data proposed by Papalexiou *et al.* (2013). The data division fulfills the condition of the availability of at least a 30-year record customary in the climate community (Arguez & Vose 2011). Individual data length for both the periods is sufficient to obtain a robust representation of the spatial pattern of the tails of the probability distribution of daily precipitation data over India.

Threshold-based approach for fitting probability distributions

In this paper, we have adopted the threshold-based approach (i.e., AES) for fitting probability distributions to the tail part of the probability distribution of non-zero daily precipitation data proposed by Papalexiou *et al.* (2013). Since the investigation revolves around the tail behavior, it is essential first to define the part of the probability distribution known as 'tail'. The demarcation of the tail of the empirical distribution for daily precipitation data by optimally selecting the threshold is a vital and crucial step in this approach. After demarcating the tail of the empirical distribution, fitting a theoretical probability distribution function to the daily precipitation data in the tail part can be accomplished by minimizing the difference between empirical and theoretical distributions.

Defining tail of empirical distribution of daily precipitation

The upper or 'right' part of the empirical probability distribution function for non-zero rainfall is referred as the 'tail'. The choice of a threshold needed for defining a tail is recognized as a difficult and open problem of debate to date. Hence, to avoid a priori selection of the threshold, we defined samples using the AES method in the present study. We choose a value x_L as a threshold such that the number of extreme precipitation events above it equals the number of years of record N (Cunnane 1973; Ben-Zvi 2009). N largest daily values of the record are preferred over each year's maximum value as the latter results in the distorted tail (Papalexiou & Koutsoyiannis 2013; Papalexiou *et al.* 2013).

The total number of non-zero daily precipitation values at a station can be computed using $n = (1 - p_0)n_d N$, where $n_d = 365.25$ is the average number of days in a year, and p_0 represents the probability of dry day. The empirical probability of exceedance $\bar{F}_N(x_i)$ is defined according to the

Weibull plotting position formula (Weibull 1939; Makkonen 2006) at each station having N-year record, and n number of non-zero precipitation values is defined as follows:

$$\bar{F}_N(x_i) = 1 - \frac{r(x_i)}{n+1} \tag{4}$$

where $r(x_i)$ is the rank of the precipitation equal to x_i in the ordered sample as $x(1) \leq \ldots . \leq x(n)$ of the non-zero values. Thus, the empirical tail is defined by the N largest non-zero precipitation values of $\bar{F}_N(x_i)$ with $n - N + 1 \leq i \leq n$. Note that the threshold value for precipitation is given as $x_L = x_{(n-N+1)}$.

Theoretical distributions considered in this study

Four simple, popular, and frequently used theoretical distributions such as Weibull (W), Lognormal (LN), Pareto-type II (PII), and the Gamma (G) distributions are considered in this study following Papalexiou *et al.* (2013) and Papalexiou *et al.* (2018). Details on these four distributions are provided in Table S1 in the Supplementary Material. Distributions selected have two parameters: one is scale parameter ($\beta > 0$) and the other is shape parameter ($\alpha > 0$). The decision on the heaviness of tails of daily precipitation data is based on fitting four probability distributions to the precipitation data in the tail part of the empirical distribution. The distributions can be divided into subexponential and exponential–hyperexponential classes based on the estimates of the shape parameter, α. The former group comprises Pareto-type II distribution, Lognormal distribution, and Weibull distribution with $\alpha < 1$, whereas the latter group includes Gamma distribution and Weibull distribution with $\alpha > 1$.

Procedure to fit probability distributions

The theoretical distributions are fitted to the precipitation records in the tail part of the empirical distribution by minimizing a PRMSE norm (which is an objective function) as given in Equation (5) (Papalexiou *et al.* 2013).

$$\text{PRMSE} = \frac{1}{N} \sum_{i=n-N+1}^{n} \left(\frac{\overline{F}(x_{(i)})}{\overline{F_N}(x_{(i)})} - 1 \right)^2 \tag{5}$$

The PRMSE norm is a function of the parameters β and α of the theoretical distributions. The norm is selected because (i) it is unbiased and suitable for subexponential distributions, (ii) it is easy to use and allows direct comparison of different distribution tails, and (iii) it gives equal weightage to each point in the tail, which contributes to the sum as relative errors between theoretical and empirical values (Papalexiou *et al.* 2013, 2018). In this study, the approach proposed by Papalexiou *et al.* (2013) is slightly modified by using a genetic algorithm (GA) (Goldberg 1989; Michalewicz *et al.* 1992) for parameter estimation of the distributions. A GA is a heuristic, stochastic, combinatorial, optimization technique based on the biological process of natural evolution (reproduction, crossover, and mutation). The heuristic is applied probabilistically to the discrete decision variables coded into binary strings. GA has been utilized effectively to minimize the PRMSE given in Equation (5) in two ways: (i) by fitting theoretical distribution to the entire precipitation data observed at a grid and (ii) by fitting theoretical distribution to N largest values of precipitation at a grid. Figure 1 depicts the approach to fit different probability distributions, namely Lognormal, Pareto-type II, Weibull, and Gamma distributions to the precipitation events in the tail part for both pre- and post-1970 periods (i.e., 1901–1970 and 1971–2010). Grids for which the parameters change but the distribution remains unchanged are shown in Figure 1. It can be inferred from the figures that the first approach where distribution is fitted to entire non-zero precipitation data does not adequately describe the tail (refer to the black dashed line). On the other hand, the solid red line representing fitting of the distributions only to the events in the tail part appears to describe the tail adequately.

RESULTS

This paper investigates the temporal and spatial changes in the behavior of daily extreme precipitation over India in terms of its frequency of occurrence due to the shift in the global climatic regime in the 1970s. The temporal changes are assessed between the two time periods, pre-1970 and post-1970, both at grid scale and regional scale. Further, a categorical classification of grids based on the change in average rainfall above threshold (increase or decrease in magnitude) and the change in the nature of the tails (i.e.,

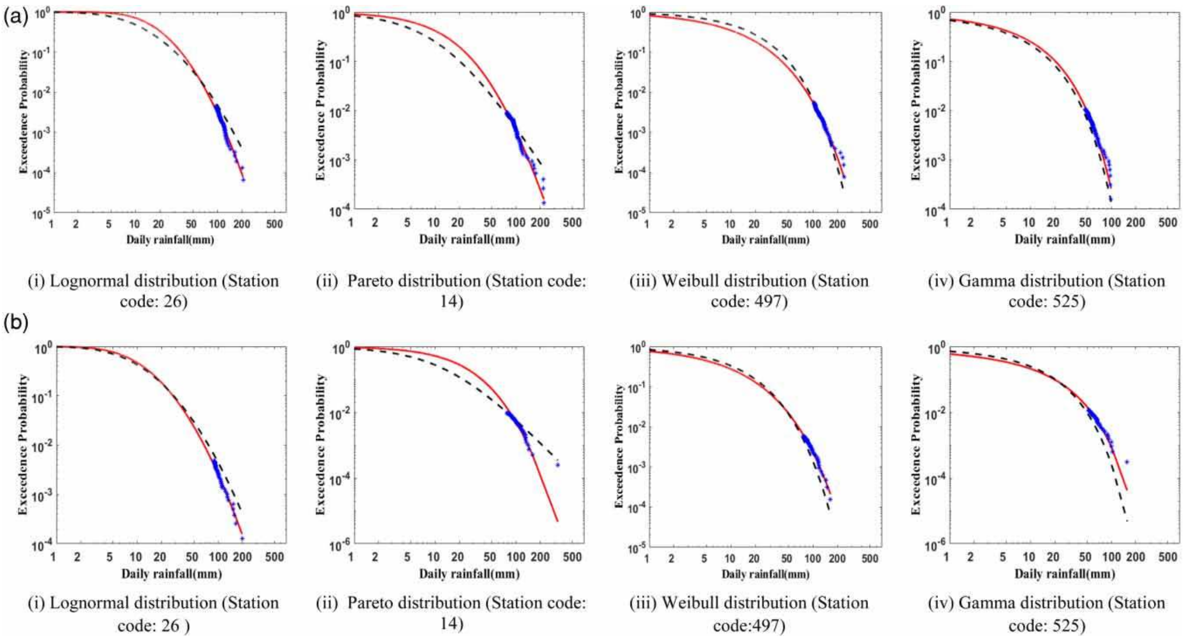

Figure 1 | PRMSE norm-based fitting approach applied to four tails, namely Lognormal, Pareto II, Weibull, and Gamma for two different periods: (a) pre-1970 (1901–1970) and (b) post-1970 (1971–2010).

from light to heavy or heavy to light) over the two time periods is also presented in this study.

Assessment of spatial and temporal changes in tail behavior of probability distribution of daily precipitation at the grid scale

In this section, the spatial and temporal changes in the behavior of tails of probability distributions of daily precipitation over India were analyzed for pre- and post-1970 periods. Following the procedure described earlier, Lognormal, Pareto-type II, Weibull, and Gamma distributions were considered to fit the non-zero daily precipitation data at 4,789 grids over India from 1901 to 1970 and 1971 to 2010, respectively. The distributions were fitted either by considering entire precipitation data at a grid or considering either 70 or 40 largest precipitation data values depicting the tail part of the distribution for the time periods of 1901–1970 and 1971–2010. Visual investigation of the fits at all 4,789 grids indicated that the fit based on the largest values in the precipitation data adequately described the tail part of the empirical distribution. This shows the advantage of the threshold-based approach proposed by Papalexiou *et al.* (2013) for fitting probability distributions to daily precipitation data, especially while analyzing the daily extreme precipitation events.

To find the best-fitted distribution of the four fitted distributions at each grid, the PRMSE norm was considered in this study. In the case of each grid, the distribution function yielding the least estimate for the PRMSE norm was declared to be the best-suited distribution for that grid. Figure 2 shows the geographical or spatial variation of best-suited distribution over India for both pre- and post-1970 periods. For the pre-1970 period, out of 4,789 grids over India, Lognormal distribution was found to be better suited for nearly 41.87% grids, followed by Pareto (32.43%), Weibull (18.56%), and Gamma (7.14%) distributions. For the post-1970s, the sequence remains the same with Lognormal as the best-fitted distribution for most grids over India. It has been observed that 45.86% of grids exhibit lognormal distribution as best-suited distribution, followed by Pareto for 32.20%, Weibull for 17.06%, and Gamma for 4.88% grids. Given these overall percentages, one may conclude that the Lognormal and Pareto-type II distributions (both heavy-tailed distributions) are the most suitable distributions for modeling the tails of probability distributions of daily precipitation data. Overall, it can be seen that there is a 4% increment and a 2.26%

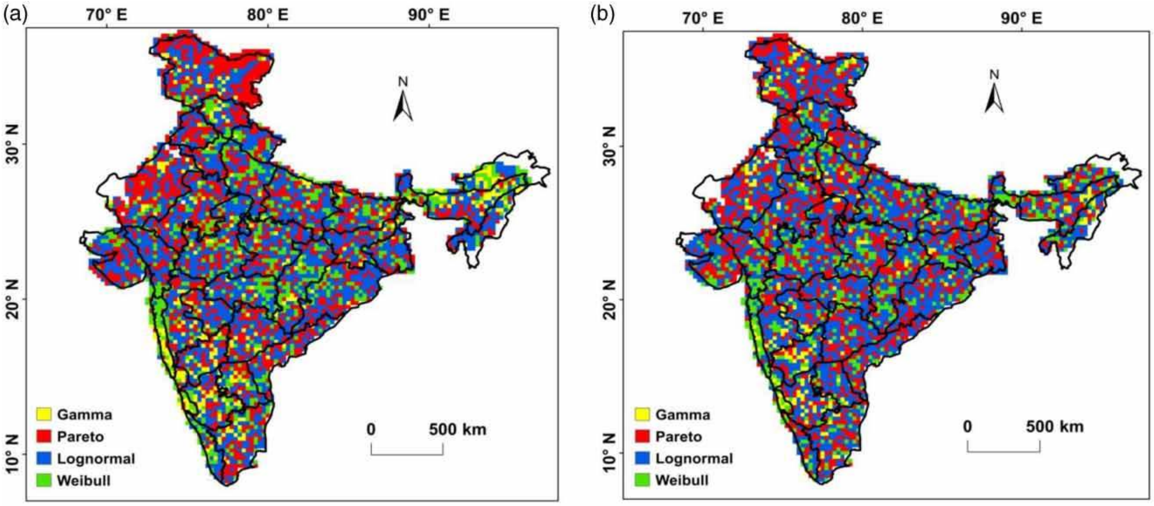

Figure 2 | Geographical locations of best-fitted distribution tail over India for two time periods: (a) pre-1970 (1907–1970) and (b) post-1970 (1971–2010). Different color coding has been used for different tail types. Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/wcc.2021.008.

decrement in the number of grids following Lognormal distribution and Gamma distribution, respectively. The Weibull distribution with a shape parameter less than 1 (which is a heavy-tailed distribution) was found to be suitable for 96.40 and 98.53% of the grids, which showed Weibull as the best-suited distribution for pre-and post-1970 periods, respectively. It can be concluded from the results that there exists a dominance of heavy-tailed distributions over light-tailed Gamma and Weibull distributions along with the increase in the tail heaviness of precipitation data over India due to climatic shifts. This indicates that extreme precipitation events in India have become more frequent.

Details on the shape (α) and scale (β) parameters of the best-suited distributions are provided in Table 1(a) and 1(b) for both pre- and post-1970 periods, respectively. The shape parameter is a scalar measure of tail behavior, and its histogram constructed based on estimates at all grids can be helpful in providing essential information about the tail heaviness. Figure 3(a) and 3(b) shows the empirical histograms of the shape parameters of four distributions considered in this study for pre- and post-1970 periods. Modal values of the histograms represent the most probable values of the shape parameters for each of the distributions. For Pareto-type II distribution, the modes were observed as 0.19 and 0.176 for the pre- and post-1970 periods. Low modal values for Pareto distribution imply the nonexistence of statistical moments for higher orders, i.e., greater than 5.26 and 5.88 (Papalexiou *et al.* 2013). The mode value of the shape parameter for Lognormal distribution was about 1.1 for both pre- and post-1970 periods. In the case of Weibull distribution, the modes of the histograms were observed to be around 0.84 for pre-1970 and 0.82 for post-1970, both implying the presence of heavier tails of the distribution as the shape parameter is less than 1. Histograms for the shape parameter of Gamma distribution show low modal values of 0.67, 0.73 for pre- and post-1970s, respectively, which indicate the presence of hyperexponential tails representing a lesser frequency of occurrence for extreme precipitation events. Histograms of shape parameters did reveal a lot about the basic nature of the tail of four distributions but to further investigate the tail relevances in describing daily precipitation, the average ranking was also considered. All four distributions were ranked in the ascending order of the PRMSE norm, i.e., the distribution yielding the least PRMSE was declared as Rank 1 distribution, while the distribution with the highest PRMSE was ranked as 4. Figure 4 illustrates the average rank of the four probability distributions for pre- and post-1970 periods. A lower average rank of a probability distribution indicates better suitability of the distribution in describing the tails of precipitation data as compared with those with higher ranks. Lognormal distribution was the best-fitted distribution with an average rank of 1.9 and 1.7 for both pre-1970 and post-1970 periods. The best-fitted distributions were ordered as Lognormal,

Table 1 | Statistical summary based on fitting of the four distributions to the tails of precipitation data for (a) pre-1970 and (b) post-1970 time periods

(a) Pre-1970 time period

	Pareto			Lognormal		
	MSE	β	α	MSE	β	α
Minimum	0.0025	0.8866	0.0000	0.0026	1.5062	0.1878
Mean	0.1014	8.1352	0.2513	0.0581	10.2748	1.2208
Maximum	0.6572	25.6076	0.6910	0.8410	24.5317	2.1150
Median	0.0438	7.6815	0.2340	0.0335	10.0345	1.1873
SD	0.1224	4.2759	0.1206	0.0718	4.7997	0.2572
Skew	1.8099	0.5574	0.4701	3.4391	0.2278	0.4291
	Weibull			**Gamma**		
	MSE	β	α	MSE	β	α
Minimum	0.0041	0.3476	0.3087	0.0037	3.4304	0.0379
Mean	0.1108	12.2075	0.9143	0.1339	17.3272	1.0737
Maximum	0.9855	31.5581	17.7532	0.4853	30.1988	7.0028
Median	0.0737	12.4927	0.8501	0.1199	17.6201	0.9060
SD	0.1221	5.0817	0.7085	0.0826	4.2256	0.7182
Skew	3.4180	−0.1809	14.4425	0.8034	−0.2213	1.9645

(b) Post-1970 time periods

	Pareto			Lognormal		
	MSE	β	α	MSE	β	α
Minimum	0.0033	0.7920	0.0033	0.0040	1.9162	0.6009
Mean	0.0892	8.1501	0.2842	0.0551	9.6792	1.3141
Maximum	0.5203	31.8910	0.7445	0.4819	24.0319	2.2519
Median	0.0487	7.5730	0.2660	0.0377	9.1881	1.2814
SD	0.0967	4.2906	0.1314	0.0535	4.6877	0.2856
Skew	1.8672	0.6598	0.4807	2.6614	0.3851	0.4276
	Weibull			**Gamma**		
	MSE	β	α	MSE	β	α
Minimum	0.0030	0.3850	0.3066	0.0040	5.1253	0.1139
Mean	0.1173	12.2138	0.8655	0.1788	18.1197	1.1743
Maximum	0.9744	30.8359	19.2964	0.6899	32.7686	10.6443
Median	0.0862	12.6215	0.8158	0.1684	18.3819	0.9731
SD	0.1111	5.1673	0.6941	0.1021	4.3883	0.8208
Skew	3.3131	−0.1866	16.2781	0.6251	−0.1592	2.6873

Pareto-type II, Weibull, and Gamma based on their ranks for both pre- and post-1970 periods. Conventionally, Gamma distribution is the most commonly used probability distribution for representing daily precipitation. However, results from this study inferred that the Gamma distribution was the worst performer for both periods. It can be noted from Figure 4

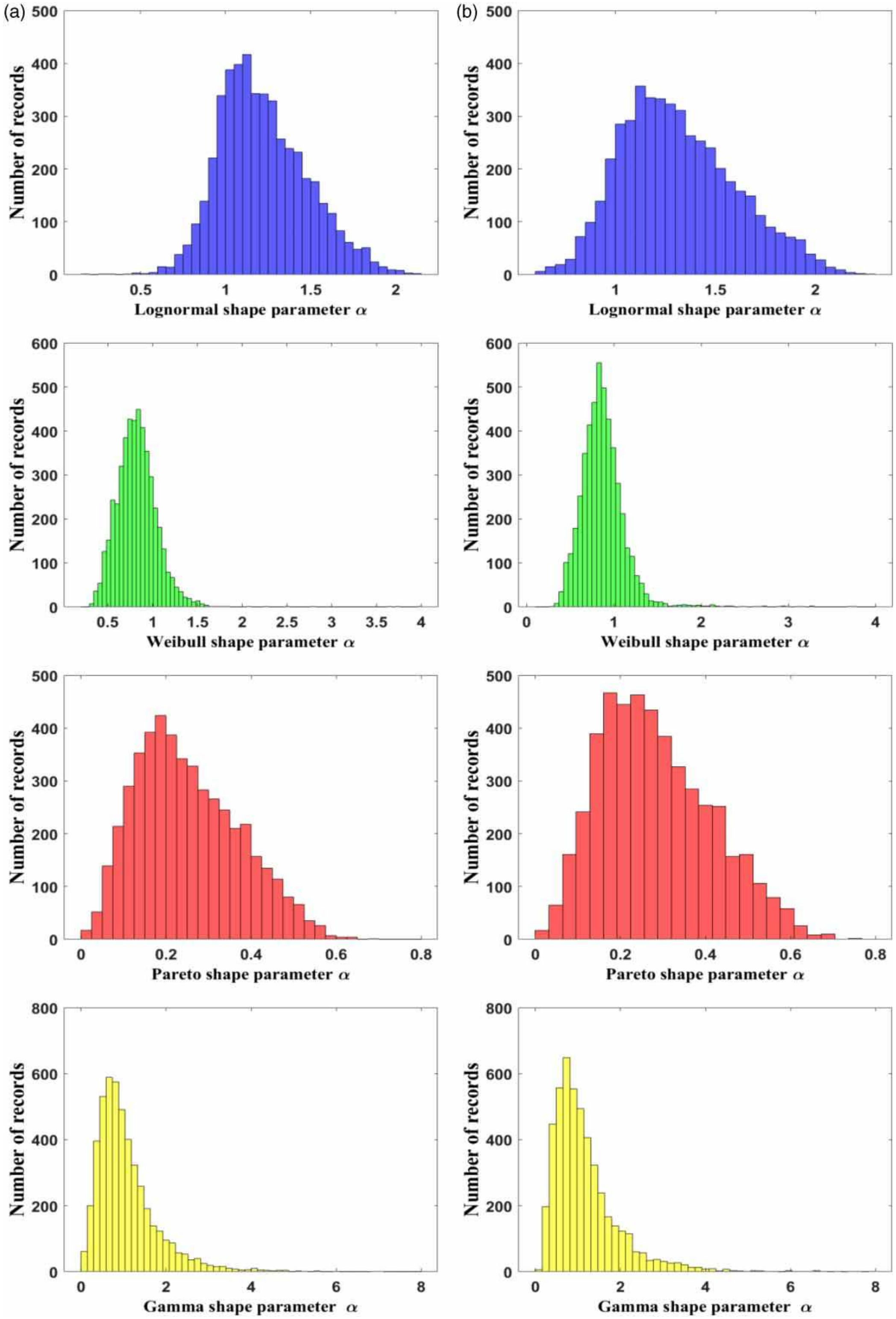

Figure 3 | Histograms of the shape parameters of four distributions fitted to all 4789 records over two time periods: (a) pre-1970 (1901–1970) and (b) post-1970 (1971–2010).

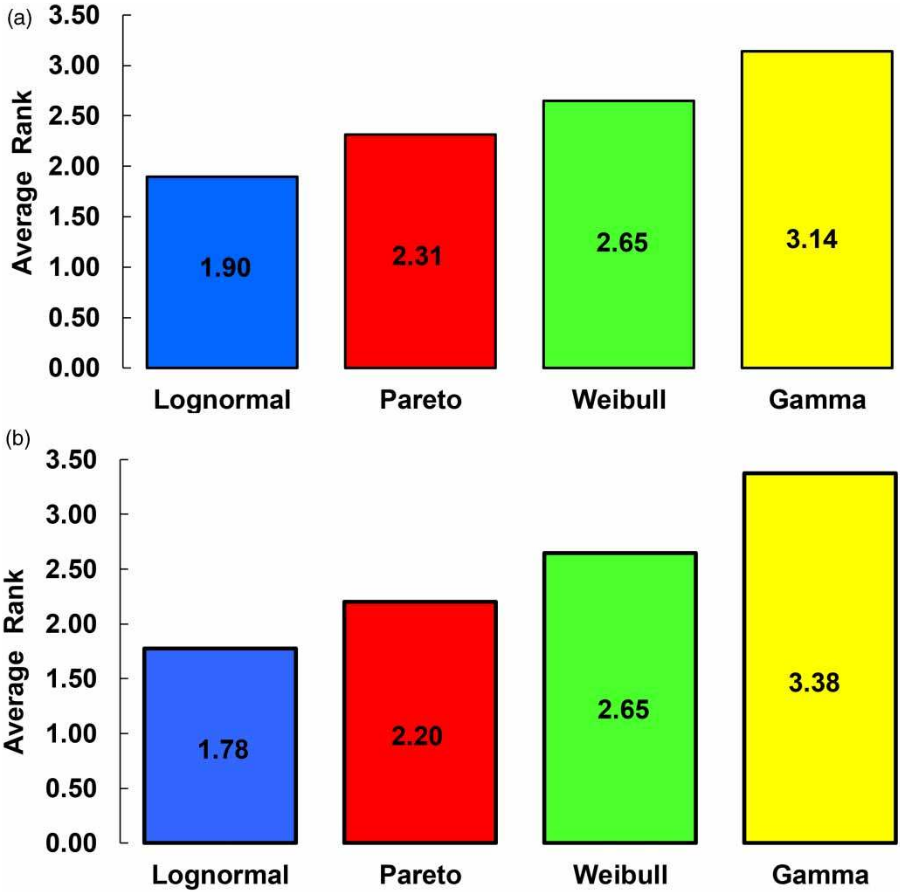

Figure 4 | Mean ranks of four distribution tails for different time periods: (a) pre-1970 (1901–1970) and (b) post-1970 (1971–2010).

that the average ranking of Lognormal and Pareto-type II distributions is decreased from 1.9 to 1.78 and 2.31 to 2.2 over the pre- and post-1970 periods, respectively. This shows that the gridded daily precipitation for the post-1970 period over India exhibits heavier tails than the pre-1970 period. Additionally, an increase in the tail heaviness of the distributions in the post-1970 period was also evident from the increase in the average rank of Gamma distribution.

Another method adopted for assessing the temporal change in the tails of probability distributions during the pre- and post-1970 periods is achieved by comparing the tails of the distributions in couples or pairs. Various pairs of distributions considered in this study are 'Lognormal vs. Pareto', 'Pareto vs. Weibull', 'Pareto vs. Gamma', 'Lognormal vs. Weibull', 'Lognormal vs. Gamma', and 'Weibull vs. Gamma'. The best-fitted distribution among the pair (any two distributions) was selected based on the PRMSE norm for each grid. The distribution with a lesser PRMSE value was considered as the best fit. Figure 5 illustrates the comparison between two probability distributions in pairs for pre- and post-1970 periods. The figure presents the percentage of grids found suitable for each probability distribution compared in pairs. It can be deduced from the figure that Lognormal distribution (which is a heavy-tailed distribution) fits the extreme daily precipitation data for 58.66 and 60.51% grids in the pre- and post-1970 periods when compared with Pareto-type II distribution. Further, the Lognormal distribution was found to be better suited than Weibull and Gamma distributions for both periods. It can be noted that the percentage of grids where daily precipitation is well represented by the Lognormal distribution against the Weibull and Gamma distributions has increased from 70.49 to 74.23% and 81.23 to 87.7%, respectively, over the pre- and post-1970 periods. This indicates that the probability distributions of daily precipitation in the post-1970 period exhibit heavier tails than the pre-1970

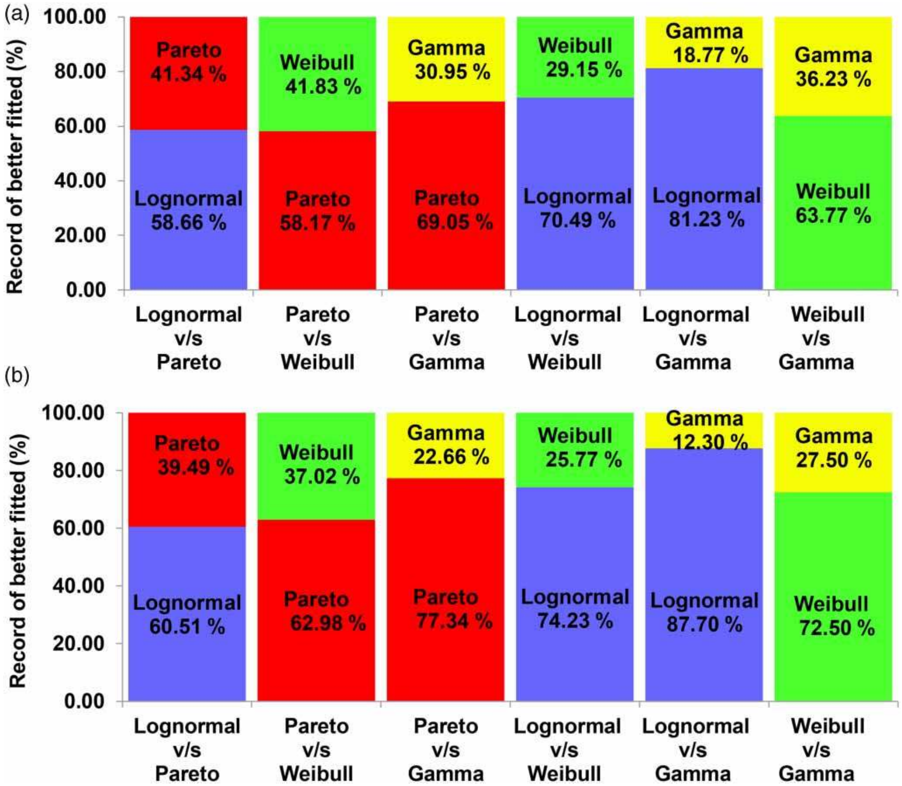

Figure 5 | PRMSE norm-based comparison of the fitted tails in couples for two time periods: (a) pre-1970 (1901–1970) and (b) post-1970 (1971–2010).

period. We have also compared Pareto-type II distribution with Lognormal, Gamma, and Weibull distributions, and it emerged as the second best-fitted distribution after Lognormal distribution. Similar to Lognormal distribution, Pareto-type II distribution was inferred as a better fitting distribution against Gamma and Weibull distributions for both pre- and post-1970 periods. The analysis revealed that the percentage of grids where the tails of probability distributions of daily precipitation are better fitted by the Pareto-type II distribution against the Weibull and Gamma distributions has increased from 58.17 to 62.98% and 69.05 to 77.34% over the periods. Among the Weibull and Gamma distributions, the Weibull distribution was better suited for describing the tails of daily precipitation data over India in both pre- and post-1970 periods. Interestingly, a heavier tailed distribution was better fitted in each case during both periods. These findings highlight that the heavier tailed distributions should be preferred over their counterparts while representing the tails of daily precipitation data over India.

We have also investigated the existence of any geographical/spatial pattern of best-suited distributions over India. The maps shown in Figure 2 illustrate the spatial distribution of best-fitted distributions for pre- and post-1970 periods. These maps do not unveil any regular patterns; instead, they seem to follow a random spatial variation. Hence, to reveal some meaningful conclusions, we categorized the best-suited distributions into either subexponential or exponential–hyperexponential classes based on the estimates of the shape parameter (α), following El Adlouni *et al.* (2008) and Papalexiou *et al.* (2013). The subexponential class includes Pareto-type II distribution, Lognormal distribution, and Weibull distribution with shape parameter <1, while the exponential–hyperexponential class comprises the Gamma distribution and Weibull distribution with shape parameter >1. Figure 6 represents maps showing the spatial distribution of subexponential and exponential–hyperexponential distributions over India for pre- and post-1970 periods. For the pre-1970 period, subexponential distributions were better suited for 4,415 out of 4,789 grids

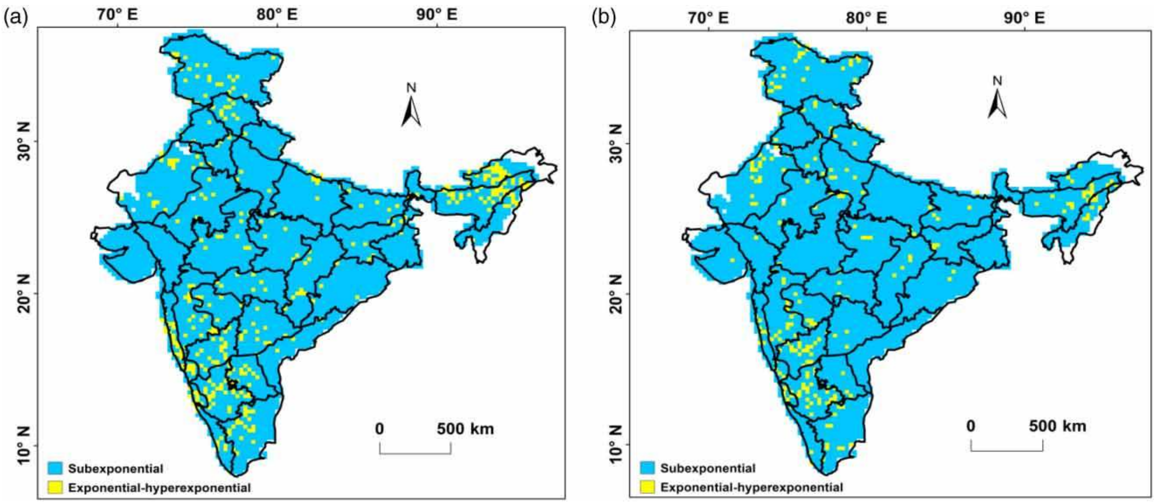

Figure 6 | Geographical variation of subexponential and exponential–hyperexponential tails over India for different periods: (a) pre-1970 (1901–1970) and (b) post-1970 (1970–2010).

(i.e., about 92.19%) over India. In comparison, the latter class was found to be applicable for merely 374 grids (i.e., only 7.81% of all the grids). Similarly, for the post-1970 period, subexponential distributions were found to be adequate to model daily precipitation data at 4,543 grids (i.e., 94.86% of all the grids while exponential–hyperexponential tails were found to be appropriate for the remaining 246 grids) (i.e., 5.14% of the grids). It can be observed from maps that the heaviness in the tails of probability distributions of daily precipitation over India has increased post-1970s climatic shift. With the dominance of heavy tails all over India, few pockets comprising lighter tails were observed in the northeast region and along the western coastal plain of the Indian Peninsula for both periods. It can be seen very well from the maps that the subexponential tails are much more dominant in the Indian region than the exponential–hyperexponential tails.

Overall, the comparison is made in terms of the difference in the percentage of the number of grids belonging to one category and the values of shape parameter (α). The maps showing the distribution and the two broad classes of tail behavior (i.e., subexponential and exponential–hyperexponential) are compared for both pre- and post-1970 periods to examine the impact of the global shift in climate regime in the 1970s on extreme precipitation. The presence of heavy tails in the daily precipitation data points to the fact that the extreme precipitation events over India are no longer rare.

Assessment of temporal changes in tail behavior of probability distribution of daily precipitation at the regional scale

Analysis at the grid scale revealed some essential inferences about the tail behavior of daily precipitation over India. However, to make the analysis more interpretable and usable at the regional scale, the temporal changes in tail behavior of the probability distribution of daily precipitation were assessed considering 34 Meteorological Subdivisions over India. Table 2 provides details about the percentage of grids having heavy or subexponential tails in each subdivision, considering the pre-1970 and post-1970 periods. Figure 7 shows the percentage of subexponential tails in each Meteorological Subdivision over India for both periods using color codes. In the case of the pre-1970 period, Saurashtra Kutch & Diu (subdivision 22) showed a complete dominance of heavy tails, followed by Gujarat (subdivision 21), Gangetic West Bengal (subdivision 6), and Orissa (subdivision 7). Nearly 22 subdivisions had more than 90% of grids showing heavy-tailed behavior. For the post-1970 period, Sub Him West Bengal and Sikkim (subdivision 5), Coastal Andhra Pradesh (subdivision 28), and Gangetic West Bengal (subdivision 6) have 100% heavy tails. In the post-1970s, 30 subdivisions were found to have more than 90% of grids showing heavy-tail behavior.

The maps showing the percentage of heavy tails in each subdivision for pre- and post-1970s were compared

Table 2 | Summary of the percentage of the grids having heavy tails within each Meteorological Subdivision for the pre- and post-1970 records

		Percentage of grids having heavy tails	
ID	Metrological region	Pre-1970 (1901–1970)	Post-1970 (1971–2010)
2	Arunachal Pradesh	68.54	95.51
3	Assam and Meghalaya	77.71	90.36
4	Naga Mani Mizo and Tripura	88.89	90.00
5	Sub Him W Bengal Sikkim	96.67	100.00
6	Gangetic West Bengal	98.92	100.00
7	Orissa	98.15	99.07
8	Jharkhand	93.86	97.37
9	Bihar	94.90	96.82
10	East Uttar Pradesh	95.48	99.10
11	West Uttar Pradesh	98.05	94.81
12	Uttaranchal	97.59	93.98
13	Haryana Chandigarh and Delhi	93.20	94.20
14	Punjab	96.67	97.78
15	Himachal Pradesh	87.64	93.18
16	Jammu and Kashmir	93.83	93.09
17	West Rajasthan	93.18	92.80
18	East Rajasthan	97.09	98.06
19	West Madhya Pradesh	95.95	98.38
20	East Madhya Pradesh	97.45	97.45
21	Gujarat	99.19	97.58
22	Saurashtra Kutch and Diu	100.00	99.36
23	Konkan and Goa	71.43	90.00
24	Madhya Maharastra	87.50	88.82
25	Marathwada	93.26	92.13
26	Vidarbha	94.89	97.08
27	Chhatisgarh	95.00	97.78
28	Coastal Andhra Pradesh	96.80	100.00
29	Telangana	92.67	98.00
30	Rayalaseema	84.44	95.56
31	Tamil Nadu and Pondicherry	88.57	93.71
32	Coastal Karnataka	58.97	74.36
33	North Interior Karnataka	82.69	81.73
34	South Interior Karnataka	78.05	78.05
35	Kerala	86.15	95.38

to find the change in the number of grids comprising heavy tails. The changes in the percentage of heavy tails over time for each subdivision are presented in Figure 7(c). In nearly 23 out of 34 subdivisions, the percentage of grids with heavy tails is observed to be increased over time. Out of those 23, nine subdivisions, namely Arunachal Pradesh, Assam and Meghalaya, Himachal Pradesh, Konkan and Goa, Telangana, Tamil Nadu, and Pondicherry, Rayalaseema, Coastal Karnataka, and Kerala, showed an increase of about 5% or above. The substantially higher percentage of grids exhibiting heavy tails post-1970 compared with that of pre-1970 might be a possible consequence of climate change and global climatic shift in the 1970s.

Assessment of temporal changes in magnitude and frequency of extreme precipitation over India

The classification of grids exhibiting the severity in terms of increase in the magnitude and frequency of extreme precipitation events due to the climatic shift is achieved by considering the combined effect of change in average precipitation above threshold and change in the tail behavior. The rainfall values above the threshold are the ones that belong to the tail. We considered the average of these values at each grid, which served as an indicator of the magnitude of the extreme precipitation. The average rainfall values above the threshold vary from a minimum value of 25.19–287.51 mm for the pre-1970 period. On the other hand, average rainfall values above the threshold range from 39.65 to 646.71 mm. Eight categories of severity were proposed by considering an increase or decrease of average rainfall above threshold and change in the nature of tail over the period from pre-1970 to post-1970. Table 3 describes the categories along with the number and percentage of grids falling in them. The categories are ranked from 1 to 8, with 1 being the most severe case and 8 being the least severe case. Figure 8 shows the spatial pattern of grids belonging to each category. Figure 8 and Table 3 show that nearly 64.5% of grids belong to Category 1, representing the most severe case. Category 2 comprises 5.42% grids where an increase in the magnitude of extreme precipitation and transition in the tail behavior from light to heavy were

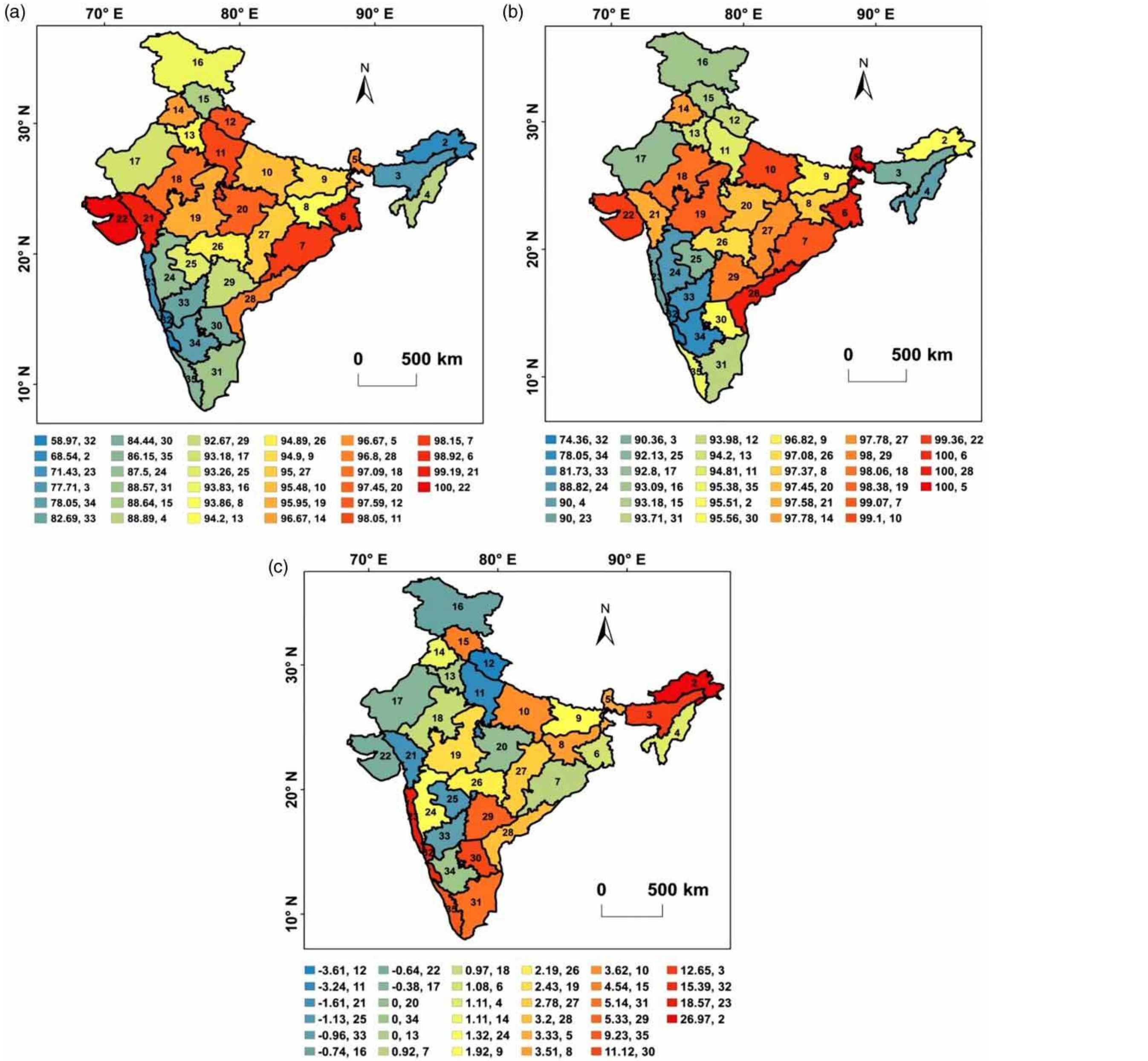

Figure 7 | Geographical variation of the percentage of subexponential tails in each Meteorological Subdivision over India for different periods: (a) pre-1970 (1901–1970) and (b) post-1970 (1970–2010). Further, changes in the percentage of grids showing heavy tails due to climate shift in the 1970s are presented in subfigure (c).

observed. Category 3 includes nearly 23.34% grids where the magnitude of extreme precipitation has decreased with heavy-tailed behavior during pre- and post-1970. Overall, most of the grids in India show a tendency of transition towards heavier tails along with an increase in the magnitude of extreme precipitation.

SUMMARY AND CONCLUSIONS

In this paper, we have analyzed the temporal and spatial changes in the tail behavior of daily precipitation over India from pre- to post-1970 time periods as per the global climatic shift. The tail behavior of precipitation data is

Table 3 | Eight categories of severity proposed by considering an increase or decrease of average rainfall above the threshold and change in the nature of tail over the period from pre-1970 to post-1970

Category (severity decreases top to bottom)	Change in average rainfall above the threshold from pre-1970 to post-1970	Change in tail type from pre-1970 to post-1970	Number of grids (out of 4789)	Percentage of grids (%)
Category 1	Increases	Heavy to heavy	3093	64.58
Category 2	Increases	Light to heavy	260	5.42
Category 3	Decreases	Heavy to heavy	1118	23.34
Category 4	Decreases	Light to heavy	72	1.50
Category 5	Increases	Heavy to light	98	2.04
Category 6	Decreases	Heavy to light	106	2.21
Category 7	Increases	Light to light	28	0.58
Category 8	Decreases	Light to light	14	0.29

assessed by identifying the best-fitted distribution out of four theoretical distributions to the sample obtained by the AES approach (e.g., Pareto-type II, Lognormal, Weibull, and Gamma distributions) based on the PRMSE norm. The approach is found to be easy to use and effective in diagnosing the tail behavior of daily precipitation data. Maps showing the geographical variation in the percentage of best-fitted subexponential tails over 34 Meteorological Subdivisions in India are given in this study. Also, the categorical classification of grids in terms of severity by considering the combined effect of an increase or decrease in average rainfall above threshold and change in the nature of tail over the period from pre-1970 to post-1970. Results from this study emphasize the importance of heavy-tailed distributions for reliable estimation of the frequency of extreme precipitation events in India. Important highlights from this study are as follows.

(i) Lognormal and Pareto-type II distributions (both heavy-tailed distributions) are found to be better suited for daily precipitation over India for both pre- and post-1970 periods. It can be concluded from the results that there exists a dominance of heavy-tailed distributions over light-tailed Gamma and Weibull distributions along with the increase in the tail heaviness of precipitation data over India due to climatic shifts. This directs us to the fact that the extreme precipitation events in India have become more frequent during both the pre- and post-1970 periods.
(ii) Gamma distribution, in general, underestimates the frequency and magnitude of extreme events. Hence, the distribution should not be considered for modeling the extreme precipitation events over India.
(iii) Histograms of shape parameters of the four probability distributions revealed that the tails of daily precipitation data have become heavier from pre- to post-1970 periods.
(iv) Heavy-tailed distributions can describe the observed precipitation extremes more effectively than light-tailed distributions. About 92.19% of the records in the pre-1970s and 94.86% in the post-1970s are better characterized by subexponential tails. Exponential–hyperexponential tails are found to be better suited for only 7.81 and 5.14% of records for the pre- and post-1970 periods. It can be seen that increasing trends of heavy tails persist in the later period, indicating a rising trend of more frequent and 'severe events' of precipitation.
(v) Twenty-three Meteorological Subdivisions in India show an increase in the percentage of heavy tails in the post-1970s compared with the pre-1970s. Further, nine subdivisions out of those 23, namely Arunachal Pradesh, Assam and Meghalaya, Himachal Pradesh, Konkan and Goa, Telangana, Tamil Nadu, and Pondicherry, Rayalaseema, Coastal Karnataka, and Kerala, showed a substantial increase in the percentage of grids exhibiting heavy tails.
(vi) Eight categories of severity are proposed by considering an increase or decrease in average rainfall above threshold and change in the nature of tail over the period from pre-1970 to post-1970. Nearly 70% of grids in India belong to Category 1 and Category 2

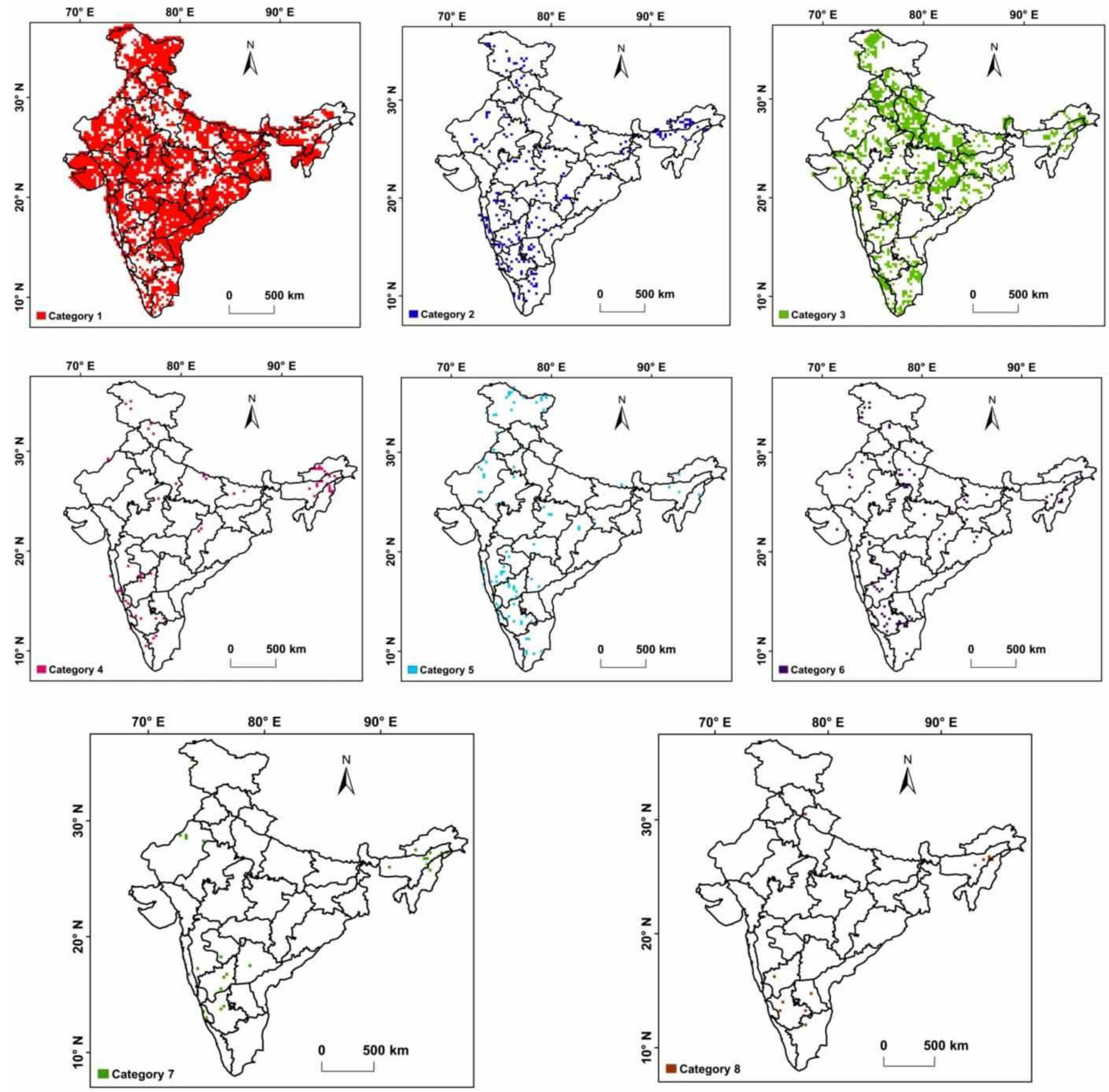

Figure 8 | Spatial pattern of grids belonging to eight categories describing severity in terms of change in increase or decrease of average extreme rainfall above threshold and change in the nature of tail over the period from pre-1970 to post-1970.

which are deemed to indicate severe/critical categories in terms of increase in the magnitude of extreme precipitation and the presence of heavier tails over pre- to post-1970 periods.

An important inference from this analysis is that the frequency and the magnitude of extreme precipitation events have generally been undervalued in the past. The use of light-tailed distributions for modeling daily precipitation can lead to a serious underestimation of the frequency and the magnitude of design extreme precipitation, which is highly undesirable for the design of water control structures. It can be noted that the results obtained from the present study are dependent on the length of precipitation records (e.g., Arguez & Vose 2011; Cavanaugh *et al.* 2015) available at each grid and the presence of serial dependence among the peaks/extreme precipitation events selected using the AES approach (e.g., Koutsoyiannis 2008). Extended research is underway to alleviate the limitation of serial dependence among the selected extreme precipitation events in the AES by exploring the strategies that can form a sample with independent events (e.g., Adams *et al.* 1986).

ACKNOWLEDGEMENTS

The authors express their gratitude to the Editor and anonymous reviewers for the constructive reviews that improved the quality of the work. The authors are also thankful to the IMD, which provided the gridded rainfall data for the study.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

REFERENCES

Adams, B. J., Fraser, H. G., Howard, C. D. & Hanafy, M. S. 1986 Meteorological data analysis for drainage system design. *Journal of Environmental Engineering* **112**, 827–848.

Ajayamohan, R. & Suryachandra, R. 2008 Indian Ocean dipole modulates the number of extreme rainfall events over India in a warming environment. *Journal of Meteorological Society of Japan* **86**, 245–252. doi:10.2151/jmsj.86.245.

Arguez, A. & Vose, R. S. 2011 The definition of the standard WMO climate normal: the key to deriving alternative climate normals. *Bulletin of the American Meteorological Society* **92** (6), 699–704. doi:10.1175/2010BAMS2955.1.

Balkema, A. & de Haan, L. 1974 Residual life time at great age. *Annals of Probability* **2** (5), 792–804.

Beguería, S. 2005 Uncertainties in partial duration series modelling of extremes related to the choice of the threshold value. *Journal of Hydrology* **303**, 215–230. https://doi.org/10.1016/j.jhydrol.2004.07.015.

Ben-Zvi, A. 2009 Rainfall intensity–duration–frequency relationships derived from large partial duration series. *Journal of Hydrology* **367**, 104–114. doi:10.1016/j.jhydrol.2009.01.007.

Beskow, S., Caldeira, T. L., de Mello, C. R., Faria, L. C. & Guedes, H. A. S. 2015 Multiparameter probability distributions for heavy rainfall modeling in extreme southern Brazil. *Journal of Hydrology: Regional Studies* **4**, 123–133.

Bingham, N. H., Goldie, C. M. & Teugels, J. L. 1987 *Regular Variation*. Cambridge University Press, Cambridge.

Biondi, F., Gershunov, A. & Cayan, D. R. 2001 North Pacific decadal climate variability since 1661. *Journal of Climate* **14**, 5–10. https:// doi.org/10.1175/1520-0442(2001)014,0005:NPDCVS.2.0.CO.

Bisht, D. S., Chatterjee, C., Raghuwanshi, N. S. & Sridhar, V. 2018a Spatio-temporal trends of rainfall across Indian river basins. *Theoretical and Applied Climatology* **132**, 419–436. https://doi.org/10.1007/s00704-017-2095-8.

Bisht, D. S., Chatterjee, C., Raghuwanshi, N. S. & Sridhar, V. 2018b An analysis of precipitation climatology over Indian urban agglomeration. *Theoretical and Applied Climatology* **133**, 421–436. https://doi.org/10.1007/s00704-017-2200-z.

Caires, S. 2009 *A Comparative Simulation Study of the Annual Maxima and the Peaks-Over-Threshold Methods*. (Deltares Report 1200264-002 for Rijkswaterstaat, Waterdienst).

Cavanaugh, N. R., Gershunov, A., Panorska, A. K. & Kozubowski, T. J. 2015 The probability distribution of intense daily precipitation. *Geophysical Research Letters* **42**, 1560–1567. doi:10.1002/2015GL063238.

Chavan, S. R. & Srinivas, V. V. 2021 Evaluation of three approaches to probable maximum precipitation estimation: a study on two Indian river basins. *Theoretical and Applied Climatology*. https://doi.org/10.1007/s00704-021-03557-5.

Chen, J. & Brissette, F. P. 2014 Stochastic generation of daily precipitation amounts: review and evaluation of different models. *Climate Research* **59**, 189–206. doi:10.3354/cr01214.

Chen, S., Yu, B. & Chen, W. 2015 An interdecadal change in the influence of the spring Arctic oscillation on the subsequent ENSO around the early 1970s. *Climate Dynamics* **44**, 1109–1126. https://doi.org/10.1007/s00382-014-2152-2.

Chistyakov, V. P. 1964 A theorem on sums of independent positive random variables and its applications to branching process. *Theory of Probability and Its Applications* **9**, 640–648.

Chow, V. T. 1964 *Handbook of Applied Hydrology*. McGraw Hill, New York.

Chowdary, J. S., Gnanaseelan, C., Vaid, B. H. & Salvekar, P. S. 2006 Changing trends in the tropical Indian ocean SST during La Nina years. *Geophysical Research Letters* **33** (18), L18610.

Cirillo, P. 2013 Are your data really Pareto distributed? *Physica A: Statistical Mechanics and its Applications* **392** (23), 5947–5962.

Coles, G. S. 2001 *An Introduction to Statistical Modeling of Extreme Values*. Springer, London.

Coles, S., Pericchi, L. R. & Sisson, S. 2003 A fully probabilistic approach to extreme rainfall modeling. *Journal of Hydrology* **273** (1–4), 35–50. https://doi.org/10.1016/S0022-1694(02)00353-0.

Cooke, R. M., Nieboer, D. & Misiewicz, J. 2014 *Fat-Tailed Distributions: Data, Diagnostics, and Dependence*. Mathematical Models and Methods in Reliability, Set No. 1. Wiley-ISTE.

Cunnane, C. 1973 A particular comparison of annual maxima and partial duration series methods of flood frequency prediction.

Journal of Hydrology **18**, 257–271. doi:10.1016/0022-1694(73)90051-6.

Dai, T., Dong, W., Guo, Y., Hong, T., Ji, D., Yang, S., Tian, D. & Xiaohangwen, Z. X. A. 2018 Understanding the abrupt climate change in the mid-1970s from a phase-space transform perspective. *Journal of Applied Meteorology and Climatology* **57**, 2551–2560. https://doi.org/10.1175/JAMC-D-17-0345.1.

Dash, S. & Maity, R. 2019 Temporal evolution of precipitation-based climate change indices across India: contrast between pre- and post-1975 features. *Theoretical and Applied Climatology*. https://doi.org/10.1007/s00704-019-02923-8.

Dash, S., Kulkarni, M. A., Mohanty, U. & Prasad, K. 2009 Changes in the characteristics of rain events in India. *Journal of Geophysical Research: Atmosphere* **114**, D10.

Deidda, R. 2010 A multiple threshold method for fitting the generalized Pareto distribution to rainfall time series. *Hydrology and Earth System Sciences* **14** (12), 2559–2575.

De Michele, C. 2019 Advances in deriving the exact distribution of maximum annual daily precipitation. *Water* **11**. https://doi.org/10.3390/w11112322.

De Michele, C. & Avanzi, F. 2018 Superstatistical distribution of daily precipitation extremes: a worldwide assessment. *Scientific Reports* **8**, 1–11. https://doi.org/10.1038/ s41598-018-31838-z.

El Adlouni, S., Bobée, B. & Ouarda, T. B. M. J. 2008 On the tails of extreme event distributions in hydrology. *Journal of Hydrology* **355**, 16–33. doi:10.1016/J.JHYDROL.2008.02.011.

Embrechts, P. & Goldie, C. M. 1980 On closure and factorization properties of subexponential and related distributions. *Journal of the Australian Mathematical Society* **29** (2), 243–256.

Embrechts, P. & Goldie, C. M. 1982 On convolution tails. *Stochastic Processes and their Applications* **13**, 263–278.

Embrechts, P., Klüppelberg, C. & Thomas, M. 1997 *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin-Heidelberg.

Feller, W. 1971 *An Introduction to Probability Theory and its Application, Vol. II*. John Wiley & Sons, New York.

Fisher, R. A. & Tippett, L. H. C. 1928 Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society* **24**, 180–190. https://doi.org/10.1017/S0305004100015681.

Foss, S., Korshunov, D. & Zachary, S. 2013 *An Introduction to Heavy-Tailed and Subexponential Distributions*. Springer-Verlag, New York. doi:10.1007/978-1-4419-9473-8.

Ghosh, S. & Resnick, S. 2010 A discussion on mean excess plots. *Stochastic Processes and Their Applications* **120** (8), 1492–1517. https://doi.org/10.1016/j.spa.2010.04.002.

Ghosh, S., Das, D., Kao, S. C. & Ganguly, A. R. 2012 Lack of uniform trends but increasing spatial variability in observed Indian rainfall extremes. *Nature Climate Change* **2** (2), 86–91. doi:10.1038/nclimate1327.

Ghosh, S., Vittal, H., Sharma, T., Karmakar, S., Kasiviswanathan, K. S., Dhanesh, Y., Sudheer, K. P. & Gunthe, S. S. 2016 Indian summer monsoon rainfall: implications of contrasting trends in the spatial variability of means and extremes. *PLoS ONE* **11** (7). doi:10.1371/journal.pone.0158670.

Giorgi, F., Im, E.-S., Coppola, E., Diffenbaugh, N. S., Gao, X. J., Mariotti, L. & Shi, Y. 2011 Higher hydroclimatic intensity with global warming. *Journal of Climate* **24**, 5309–5324. doi:10.1175/4092011JCLI3979.1.

Gnedenko, B. 1943 Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics* **44** (3), 423–453. https://doi.org/10.2307/1968974.

Goldberg, D. E. 1989 *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, Reading, MA, New York, p. 126.

Goldie, C. M. & Klüppelberg, C. 1998 Subexponential distributions. In: *A Practical Guide to Heavy Tails: Statistical Techniques and Applications* (R. J. Adler, R. E. Feldman & M. S. Taqqu, eds). Springer, Birkhäuser-Boston, pp. 435–459.

Goswami, B. N., Venugopal, V., Sengupta, D., Madhusoodanan, M. S. & Xavier, P. K. 2006 Increasing trend of extreme rain events over India in a warming environment. *Science* **314**, 1442–1445. doi:10.1126/science.1132027.

Graham, N. E. 1994 Decadal-scale climate variability in the tropical and North Pacific during the 1970 and 1980s: observations and model results. *Climate Dynamics* **10**, 135–162.

Groisman, P. Y., Karl, T. R., Easterling, D. R., Knight, R. W., Jamason, P. F., Hennessy, K. J., Suppiah, R., Page, C. M., Wibig, J., Fortuniak, K., Razuvaev, V. N., Douglas, A., Førland, E. & Zhai, P. 1999 Changes in the probability of heavy precipitation: important indicators of climatic change. In: *Weather and Climate Extremes* (T. R. Karl, N. Nicholls & A. Ghazi, eds). Springer, Dordrecht, pp. 243–283. https://doi.org/10.1007/978-94-015-9265-9_15.

Guhathakurta, P. & Rajeevan, M. 2008 Trends in the rainfall pattern over India. *International Journal of Climatology* **28**, 1453–1469. doi:10.1002/joc.1640.

Gupta, S. K. 2011 *Modern Hydrology and Sustainable Water Development*. John Wiley & Sons. https://doi.org/10.1002/9781444323962.

Hill, B. M. 1975 A simple general approach to inference about the tail of a distribution. *The Annals of Statistics* **3**, 1163–1174.

Hill, S. A. 2019 A measure for characterizing heavy-tailed networks. arXiv:1907.04808v1.

Hsu, P.-C. 2016 Global monsoon in a changing climate. In: *The Monsoons and Climate Change: Observations and Modeling* (L. M. Véspoli de Carvalho & C. Jones, eds). Springer Climate, Switzerland, pp. 7–24.

Huang, J., Yu, H., Dai, A., Wei, Y. & Kang, L. 2017a Drylands face potential threat under 28°C global warming target. *Nature Climate Change* **7**, 417–422. https://doi.org/10.1038/nclimate3275.

Huang, J., Li, Y., Fu, C., Chen, F., Fu, Q., Dai, A., Shinoda, M., Ma, Z., Guo, W., Li, Z., Zhang, L., Liu, Y., Yu, H., He, Y., Xie, Y., Guan, X., Ji, M., Lin, L., Wang, S., Yan, H. & Wang, G. 2017b Dryland climate change: recent progress and challenges. *Reviews of Geophysics* **55** (3), 719–778.

IPCC 2013 Climate Change 2013: The Physical Science Basis. In: *Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. (T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex & P. M. Midgley, eds.), Cambridge, UK and New York, NY, USA: Cambridge University Press.

Jacob, D. & Hagemann, S. 2007 Intensification of the hydrological cycle: an important signal of climate change. In: *Global Change: Enough Water for all?* (J. L. Lozan, ed.). Wissenschaftliche Auswertungen, Hamburg, pp. 170–173

Jacques-coper, M. & Garreaud, R. D. 2015 Characterization of the 1970s climate shift in South America. *International Journal of Climatology* **35**, 2164–2179. https://doi.org/10.1002/joc.4120.

Jenkinson, A. F. 1955 The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society* **81** (348), 158–171. https://doi.org/10.1002/qj.49708134804.

Kiran, K. G. & Srinivas, V. V. 2021 A Mahalanobis distance-based automatic threshold selection method for peaks over threshold model. *Water Resources Research* **57** (1), e2020WR027534.

Klüppelberg, C. 1988 Subexponential distributions and integrated tails. *Journal of Applied Probability* **25** (01), 132–141. https://doi.org/10.2307/3214240.

Kotz, S. & Nadarajah, S. 2000 *Extreme Value Distributions: Theory and Applications*. Imperial College Press, London. https://doi.org/10.1142/p191.

Koutsoyiannis, D. 2004a Statistics of extremes and estimation of extreme rainfall, 1, theoretical investigation. *Hydrological Sciences Journal* **49**, 575– 590.

Koutsoyiannis, D. 2004b Statistics of extremes and estimation of extreme rainfall, 2, empirical investigation of long rainfall records. *Hydrological Sciences Journal* **49**, 591–610.

Koutsoyiannis, D. 2008 *Probability and Statistics for Geophysical Processes*. National Technical University of Athens. https://doi.org/10.13140/RG.2.1.2300.1849/1.

Kozubowski, T. J., Panorska, A. K., Qeadan, F., Gershunov, A. & Rominger, D. 2009 Testing exponentiality versus Pareto distribution via likelihood ratio. *Communication in Statistics- Simulation and Computation* **38**, 118–139. doi:10.1080/03610910802439121.

Kulkarni, M. A., Singh, A. & Mohanty, U. 2012 Effect of spatial correlation on regional trends in rain events over India. *Theoretical Applied Climatology* **109** (3–4), 497–505.

Langousis, A., Mamalakis, A., Puliga, M. & Deidda, R. 2016 Threshold detection for the generalized Pareto distribution: review of representative methods and application to the NOAA NCDC daily rainfall database. *Water Resources Research* **52** (4), 2659–2681. https://doi.org/10.1002/2015WR018502.

Li, C., Singh, V. P. & Mishra, A. K. 2012 Simulation of the entire range of daily precipitation using a hybrid probability distribution. *Water Resources Research* **48**, W03521. doi:10.1029/2011WR011446.

Madsen, H., Rasmussen, P. F. & Rosbjerg, D. 1997 Comparison of annual maximum series and partial duration series methods for modelling extreme hydrologic events: 1. At-site modeling. *Water Resources Research* **33** (4), 747–757.

Mailhot, A., Lachance-Cloutier, S., Talbot, G. & Favre, A. C. 2013 Regional estimates of intense rainfall based on the peak-over-threshold (POT) approach. *Journal of Hydrology* **476**, 188–199. https://doi.org/10.1016/j.jhydrol.2012.10.036.

Makkonen, L. 2006 Plotting positions in extreme value analysis. *Journal of Applied Meteorology and Climatology* **45** (2), 334–340. https://doi.org/10.1175/JAM2349.1.

Meehl, G. A., Hu, A. & Santer, B. D. 2008 The mid-1970s climate shift in the pacific and the relative roles of forced versus inherent decadal variability. *Journal of Climate* **22**, 780–792. https://doi.org/10.1175/2008JCLI2552.1.

Michalewicz, Z., Janikow, C. Z. & Krawczyk, J. B. 1992 A modified genetic algorithm for optimal control problems. *Computers and Mathematics with Applications* **23** (12), 83–94.

Mielke, J. 1973 PW another family of distributions for describing and analyzing precipitation data. *Journal of Applied Meteorology and Climatology* **12**, 275–280.

Mlyński, D., Wałęga, A., Petroselli, A., Tauro, F. & Cebulska, M. 2019 Estimating maximum daily precipitation in the upper Vistula basin, Poland. *Atmosphere (Basel)* **10**, 43. https://doi.org/10.3390/atmos10020043.

Moccia, B., Mineo, C., Ridolfi, E., Russo, F. & Napolitano, F. 2021 Probability distributions of daily rainfall extremes in Lazio and Sicily, Italy, and design rainfall inferences. *Journal of Hydrology: Regional Studies* **33**, 100771.

Mondal, A. & Mujumdar, P. P. 2015 Modeling non-stationarity in intensity, duration and frequency of extreme rainfall over India. *Journal of Hydrology* **521**, 217–231.

Nagaev, A. & Tsitsiashvili, G. 2006 Tail asymptotics of the nth convolution of super-exponential distributions. *Statistics & Probability Letters* **76** (9), 861–870.

Nerantzaki, S. & Papalexiou, S. M. 2019 Tails of extremes: advancing a graphical method and harnessing big data to assess precipitation extremes. *Advances in Water Resources* **134**. https://doi.org/10.1016/j.advwatres.2019.103448.

Nieboer, D. 2011 *Heuristics of Heavy-Tailed Distributions and the Obesity Index*. *Dissertation*, Delft University of Technology.

O'Kane, T. J., Matear, R. J., Chamberlain, M. A. & Oke, P. R. 2014 ENSO regimes and the late 1970's climate shift: the role of synoptic weather and south Pacific Ocean spiciness. *Journal of Computational Physics* **271**, 19–38.

Ouarda, T. B. M. J., Ashkar, F., Bensaid, E. & Hourani, I. 1994 Statistical distributions used in hydrology. Transformations and asymptotic properties. *Scientific Report*, Department of Mathematics, University of Moncton, New Brunswick, p 31.

Pai, D., Sridhar, L., Rajeevan, M., Sreejith, O., Satbhai, N. & Mukhopadhyay, B. 2014 Development of a new high spatial resolution (0.25×0.25) long period (1901–2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region. *Mausam* **65**, 1–18. http://www.imd.gov.in/advertisements/20170320_advt_34.

Panahi, H. 2016 Model selection test for the heavy-tailed distributions under censored samples with application in financial data. *International Journal of Financial Studies* **4** (4), 24.
Panorska, A. K., Gershunov, A. & Kozubowski, T. J. 2007 From diversity to volatility: probability of daily precipitation extremes. In: *Nonlinear Dynamics in Geophysics* (A. A. Tsonis & J. B. Elsner, eds). Springer, New York, pp. 465–484.
Papalexiou, S. M. & Koutsoyiannis, D. 2013 Battle of extreme value distributions: a global survey on extreme daily rainfall. *Water Resources Research* **49**, 187–201. doi:10.1029/2012WR012557.
Papalexiou, S. M. & Koutsoyiannis, D. 2016 A global survey on the seasonal variation of the marginal distribution of daily precipitation. *Advances in Water Resources* **94**, 131–145. https://doi.org/10.1016/J.ADVWATRES.2016.05.005.
Papalexiou, S. M., Koutsoyiannis, D. & Makropoulos, C. 2013 How extreme is extreme? An assessment of daily rainfall distribution tails. *Hydrology & Earth System Sciences* **17** (2), 851–862. https://doi.org/10.5194/hess-17-851-2013.
Papalexiou, S. M., AghaKouchak, A. & Foufoula-Georgiou, E. 2018 A diagnostic framework for understanding climatology of tails of hourly precipitation extremes in the United States. *Water Resources Research* **54**, 6725–6738. https://doi.org/10.1029/2018WR022732.
Pickands, J. I. 1975 Statistical inference using extreme order statistics. *Annals of Statistics* **3**, 119–131.
Rajeevan, M., Bhate, J. & Jaswal, A. K. 2008 Analysis of variability and trends of extreme rainfall events over India using 104 years of gridded daily rainfall data. *Geophysical Research Letters* **35**, L18707.
Roxy, M. K., Ghosh, S., Pathak, A., Athulya, R., Mujumdar, M., Murtugudde, R., Terray, P. & Rajeevan, M. 2017 A threefold rise in widespread extreme rain events over central India. *Nature Communication* **8** (1–11). https://doi.org/10.1038/s41467-017-00744-9.
Sabeerali, C. T., Rao, S. A., Ajayamohan, R. S. & Murtugudde, R. 2012 On the relationship between Indian summer monsoon withdrawal and Indo-Pacific SST anomalies before and after 1976/1977 climate shift. *Climate Dynamics* **39** (3), 841–859.
Sahana, A. S., Ghosh, S., Ganguly, A. & Murtugudde, R. 2015 Shift in Indian summer monsoon onset during shift in Indian summer monsoon onset during 1976/1977. *Environmental Research Letters* **10** (5), 054006. https:// doi.org/10.1088/1748-9326/10/5/054006.
Sarkar, S. & Maity, R. 2020 Increase in probable maximum precipitation in a changing climate over India. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2020.124806.
Serinaldi, F. & Kilsby, C. G. 2014 Rainfall extremes: toward reconciliation after the battle of distributions. *Water Resources Research* **50**, 336–352. https://doi.org/10.1002/2013WR014211.
Shastri, H., Paul, S., Ghosh, S. & Karmakar, S. 2015 Impacts of urbanization on Indian summer monsoon rainfall extremes. *Journal of Geophysical Research: Atmospheres* **120**, 495–516. doi:10.1002/2014JD022061.
Singh, J., Vittal, H., Karmakar, S., Ghosh, S. & Niyogi, D. 2016 Urbanization causes non-stationarity in Indian summer monsoon rainfall extremes. *Geophysical Research Letters* **43** (21), 11–269.
Teugel, J. L. 1975 The class of subexponential distributions. *The Annals of Probability.* doi:10.2307/2959204.
Vela, A. C. & Rodríguez, G. 2014 *Extreme Value Theory: An Application to the Peruvian Stock Market Returns*. (No. 2014-394) Documentos de Trabajo/Working Papers.
Villarini, G., Smith, J. A., Ntelekos, A. A. & Schwarz, U. 2011 Annual maximum and peaks-over-threshold analyses of daily rainfall accumulations for Austria. *Journal of Geophysical Research* **116**, D05103. doi:10.1029/2010JD015038.
Vittal, H., Karmakar, S. & Ghosh, S. 2013 Diametric changes in trends and patterns of extreme rainfall over India from pre-1950 to post-1950. *Geophysical Research Letters* **40** (12), 3253–3258.
Voitalov, I., van der Hoorn, P., van der Hofstad, R. & Krioukov, D. 2018 *Scale-Free Networks Well Done*. https://arxiv.org/abs/1811.02071.
Wang, H. 2001 The weakening of the Asian monsoon circulation after the end of 1970's. *Advances in Atmospheric Sciences* **18** (3), 376–386. https://doi.org/10.1007/BF02919316.
Wang, Y., Xu, H., Cheng, D. & Yu, C. 2018 The local asymptotic estimation for the supremum of a random walk with generalized strong subexponential summands. *Statistical Papers* **59** (1), 99–126.
Weibull, W. 1939 A statistical theory of strength of materials. *Ing. Vetensk. Akad. Handl* **151**, 1–45.
Weisheimer, A., Schaller, N., O'Reilly, C., MacLeod, D. A. & Palmer, T. 2017 Atmospheric seasonal forecasts of the twentieth century: multi-decadal variability in predictive skill of the winter North Atlantic Oscillation (NAO) and their potential value for extreme event attribution. *Quarterly Journal of Royal Meteorological Society* **143**, 917–926. https://doi.org/10.1002/qj.2976.
Werner, T. & Upper, C. 2004 Time variation in the tail behavior of bund future returns. *Journal of Futures Markets* **24** (4), 387–398. doi:10.1002/fut.10120.
Zhang, R. H., Rothstein, L. M. & Busalacchi, A. J. 1998 Origin of Upper Ocean warming and El Niño change on decadal scales in the tropical Pacific Ocean. *Nature* **391**, 879–883.
Zhou, T., Gong, D., Li, J. & Li, B. 2009 Detecting and understanding the multi-decadal variability of the East Asian Summer Monsoon–Recent progress and state of affairs. *Meteorologische Zeitschrift* **18** (4), 455–467.
Zuo, J., Ren, H.-L., Li, W. & Wang, L. 2016 Interdecadal variations in the relationship between the winter North Atlantic oscillation and temperature in south-central China. *Journal of Climate* **29**, 7477–7493. https://doi.org/10.1175/JCLI-D-15-0873.1.

First received 3 January 2021; accepted in revised form 4 May 2021. Available online 17 May 2021