



# A multistate first-order Markov model for modeling time distribution of extreme rainfall events

A. N. Rohith<sup>1,2</sup> · Margaret W. Gitau<sup>2</sup> · I. Chaubey<sup>3</sup> · K. P. Sudheer<sup>1,2,4</sup>

Received: 29 June 2020 / Revised: 12 October 2020 / Accepted: 12 November 2020 / Published online: 2 December 2020  
© The Author(s) 2020

## Abstract

The time distribution of extreme rainfall events is a significant property that governs the design of urban stormwater management structures. Accuracy in characterizing this behavior can significantly influence the design of hydraulic structures. Current methods used for this purpose either tend to be generic and hence sacrifice on accuracy or need a lot of model parameters and input data. In this study, a computationally efficient multistate first-order Markov model is proposed for use in characterizing the inherently stochastic nature of the dimensionless time distribution of extreme rainfall. The model was applied to bivariate extremes at 10 stations in India and 205 stations in the United States (US). A comprehensive performance evaluation was carried out with one-hundred stochastically generated extremes for each historically observed extreme rainfall event. The comparisons included: 1-h (15-min); 2-h (30-min); and, 3-h (45-min) peak rainfall intensities for India and (US) stations, respectively; number of first, second, third, and fourth-quartile storms; the dependence of peak rainfall intensity on total depth and duration; and, return levels and return periods of peak discharge when these extremes were applied on a hypothetical urban catchment. Results show that the model efficiently characterizes the time distribution of extremes with: Nash–Sutcliffe-Efficiency  $> 0.85$  for peak rainfall intensity and peak discharge;  $< 20\%$  error in reproducing different quartile storms; and,  $< 0.15$  error in correlation analysis at all study locations. Hence the model can be used to effectively reproduce the time distribution of extreme rainfall events, thus increasing the confidence of design of urban stormwater management structures.

**Keywords** Extreme rainfall events · Time-distribution of rainfall · First-order Markov model · Dimensionless mass curve · Return periods

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00477-020-01939-1>) contains supplementary material, which is available to authorized users.

✉ Margaret W. Gitau  
mgitau@purdue.edu

<sup>1</sup> Department of Civil Engineering, Indian Institute of Technology Madras, Chennai, Tamil Nadu, India

<sup>2</sup> Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, Indiana, USA

<sup>3</sup> College of Agriculture, Health and Natural Resources, University of Connecticut, Storrs, Connecticut, USA

<sup>4</sup> Kerala State Council for Science, Technology, and Environment, Thiruvananthapuram, Kerala, India

## 1 Introduction

Statistical analysis and characterization of extreme rainfall is a critical part of the design of various hydraulic structures related to urban stormwater management. Recent studies have endorsed the use of event-based multivariate analysis over a critical duration-based univariate design storm approach for this purpose (Park et al. 2013; Balistrocchi and Bacchi 2017; Jun et al. 2017, 2018). This is due to the flexibility that event-based analysis provides, allowing critical rainfall properties to be characterized more realistically than through critical duration-based analysis. For example, with a univariate design storm approach, the frequency analysis is carried out only on the maximum rainfall depth that occurred over a predefined critical duration. This critical duration is usually less than the actual rainfall event duration, hence only a part of the rainfall event is considered. This could misrepresent the

antecedent wetness, which may further lead to underestimation of catchment response (Berk et al. 2017). With event-based analysis, the actual duration, depth, and peak intensity of rainfall events are studied together, which results in a more realistic analysis.

In event-based analysis, an independent rainfall event is defined as a continuous stretch of rainfall in time separated by a minimum period (event separation time) of no rainfall (Joo et al. 2013; Lee and Kim 2018). Independent rainfall events possess multiple properties such as total rainfall depth (R), duration (D), peak intensity (I), and time distribution during the occurrence. The results of rainfall-runoff models are usually affected by all these properties of extremes, the association between them (Azarnivand et al. 2020), and the input timestep of rainfall pulses (Sampson et al. 2020). Multivariate frequency analysis based on copulas (Nelsen 2006) is generally adequate to characterize the extreme rainfall properties R, D, and I and the associations among them (Kao and Govindaraju 2007; Zhang and Singh 2007). However, the time distribution of a rainfall event is one of the most preeminent properties, which affects urban catchment response (Zeimet et al. 2018) and thereby the design of hydraulic structures. Hence, it is essential to understand and appropriately characterize the time distribution of rainfall within an event.

Since mid-20th century, several methods have been developed (Huff 1967; Schertzer and Lovejoy 1987; Rodríguez-Iturbe et al. 1987), improved (Entekhabi et al. 1989), or modified (Onof and Wheater 1994), to model the time behavior of the rainfall within events at a fine time scale (hourly or less). Current methods include those based on or using: point process models (Onof et al. 2000; Verhoest et al. 2010; Kossieris et al. 2018); the theory of scale invariance (Molnar and Burlando 2005; Sivakumar and Sharma 2008); empirical geometric shapes for hyetographs (Huff 1967); multivariate copula-based frequency analysis (Zhang and Singh 2007; Kao and Govindaraju 2008; Jun et al. 2017; Guo et al. 2018); and, Markovian processes (Nguyen and Rousselle 1981; Woolhiser and Osborn 1985).

Point process-based methods are some of the most robust approaches available for modeling time distribution of rainfall events (Olsson and Burlando 2002). These methods treat a rainfall event as a cluster of many small rainy cells with a random period of rain (Rodríguez-Iturbe et al. 1987; Rodríguez-Iturbe et al. 1988; Onof and Wheater 1994). However, point process models require a large number of parameters (5–7), and have shown some disagreements (Koutsoyiannis and Mamassis 2001). Scale-based methods use the inherent scale-invariant structure of rainfall (Schertzer and Lovejoy 1987; Gupta et al. 1993) with the assumption that the statistical properties of rainfall occurrences aggregated at different time scales are related

through a simple scaling function. One such approach is a multiplicative random cascade model, where the total rainfall is distributed on consecutive subdivisions (Molnar and Burlando 2005; Sivakumar and Sharma 2008). Though the assumption of scale-invariance is valid in many cases, some authors (Marani 2003; Di Baldassarre et al. 2006) have reported that this assumption is repeatedly violated at finer time scales and may not be valid in many instances.

One of the most widely used ways of representing rainfall events is through the dimensionless rainfall mass curves [DRMC, Huff (1967)]. These comprise nine probabilistic DRMC (Huff curves) at 10–90% probabilities in each quartile storm duration, combined to a total of 36 DRMC. The 50% curve is the single most representative huff curve in each quartile (Huff 1990); however, the selection of one single huff curve for practical purposes is still unclear (Bonta 2004). Four synthetic dimensionless mass curves named type I, IA, II, and III developed for use in US watersheds (US-SCS 1986) have been used in studies such as (Awadallah and Younan 2012; Ghassabi et al. 2016) to define the shape of design storms. However, these curves have been found to overestimate peak discharges (Guo and Hargadin 2009; Kimoto et al. 2011; Dullo et al. 2017). Few studies (Acreman 1990; Garcia-Guzman and Aranda-Oliver 1993) have assumed that the shape of the rainfall mass curve (RMC) follows the beta distribution and observed that the peak rainfall intensities and their frequencies were reasonably reproduced. Kottegoda et al. (2003) used a geometric distribution for daily wet and dry runs, and a beta distribution for hourly rainfall magnitudes, and observed that the results were satisfactory. Hingray et al. (2002), performed stochastic generation and disaggregation of hourly rainfall data and concluded that the stochastic methods are well suited for the purpose.

A classical approach to model storms of complex nature is by describing them as Markovian process. One way is to model the occurrence of rainfall as a two-state Markov process and depth of rainfall with a theoretical probability distribution (e.g. Nguyen and Rousselle 1981). The second approach is to model both rainfall depth and occurrence as a multi-state Markovian process by dividing the rainfall amount into several discrete states, including a state with no rain (Woolhiser and Osborn 1985). These latter authors defined rescaled dimensionless rainfall increments at ten dimensionless time steps following a first-order Markov chain. Further, they assumed that the marginal and conditional distributions of the incremental process followed the beta distribution and recommended a 13-parameter model. However, fitting such a large number of parameters can increase the uncertainty due to parameter estimation. Moreover, the marginal and conditional distributions of a multistate Markov process need a substantial amount of data (Sharma and Mehrotra 2010).

In general, current methods used to model the time distribution of rainfall events need a relatively large number of parameters and a substantial amount of observed data to obtain suitable results. Conversely, simpler, fewer parameter models and empirical dimensionless profiles may not produce satisfactory results. Regardless, comprehensive performance evaluation of existing methods is needed to determine their ability to reproduce critical properties of rainfall events. The occurrence of rainfall within an event is highly stochastic. A non-parametric multistate Markov process-based model applied to dimensionless rainfall events could provide the needed improvements. However, its efficiency greatly depends on the selection of the number of states, the correlation between conditional and marginal processes, and dependence of time distribution with other properties of rainfall events.

In this study, a computationally efficient, non-parametric first-order multistate Markov model was developed to characterize the time distribution of extreme rainfall events. The proposed approach is empirical and assumes no theoretical probability distributions for both conditional and marginal probabilities of the Markovian process. Further, a comprehensive performance evaluation was carried out to validate the proposed methodology. The new approach was tested for its ability to reproduce: peak rainfall intensities of different durations; fraction of first, second, third, and fourth quartile storms; and, the dependence structure between R and I, and D and I. These rainfall events were further run through a hypothetical urban catchment to assess the model's ability to reproduce direct runoff properties such as peak runoff and its return period.

## 2 Material and methods

### 2.1 Study area

The methodology was developed and tested on the hourly rainfall data recorded at ten cities in India (Fig. 1a). It was then applied to 15-min rainfall data from 205 rainfall stations in the US (Fig. 1b). The rainfall stations selected are located in urbanized areas and represent a wide range of climatological and geographical characteristics. The India cities were selected based on three criteria: broad geographical coverage; climatological representation; and, data availability (at least 40 years). The country falls under both tropical and subtropical climates (Beck et al. 2018) with a wide range of variability in weather and geography. Hourly rainfall data recorded using automatic weather stations at these study locations were procured from the Data Supply Center of the India Meteorological Department in Pune, India.

Fifteen-minute rainfall data for the US stations were downloaded from the National Oceanic and Atmospheric Administration, data portal (<https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ncdc:C00505>). These stations were selected based on data availability over a period of at least 40 years and their proximity to an urban area (within 5 km from the urban boundary) as identified based on the 2010 Census for the US, and data from ESRI (<https://www.arcgis.com/home/item.html?id=069b5cafe3e34a2585e24ba63cd12b9e>). The list of these stations with their associated latitude, longitude, and climate region is provided with supplementary materials (Table S1).

### 2.2 Definition of extreme rainfall events

At first, continuously recorded hourly rainfall was delineated into independent and isolated rainfall events using a 6-h inter-event time. A 6-h inter-event time is commonly used in many studies to identify independent rainfall events (Huff 1967; Kao and Govindaraju 2007). A rainfall intensity threshold of 1 mm/h was adopted to identify rainfall pulses as this is the minimum intensity required to classify precipitation as rainfall (Subramanya 1994; Deodhar 2008). All rainfall events with missing data within an event were excluded from the study so as not to misrepresent the observed time distribution. Such rainfall events were largely uncommon. Among delineated rainfall events, only extremes affect the design of water resources structures; however, the definition of a rainfall event as extreme is not straight forward (Kao and Govindaraju 2007). This is because multiple properties of rainfall events (R and I) usually govern design, and, thus, extreme rainfall events must possess extreme nature in these properties. In this study, a bivariate Peak Over Threshold (POT) approach was adopted to extract extreme rainfall events.

The bivariate POT approach used in this study defines a threshold on empirical copula ( $K_n$ ) constructed between R and I of all the rainfall events and goodness of fit test to define an optimal threshold. At first, a threshold of  $K_n = 0.95$  was defined as an initial point, and all the rainfall events above the threshold were assumed as extremes. A goodness of fit test on R, I, and D of these extremes was further carried out for various probability distributions such as Generalised Extreme Value, Generalised Pareto, Log-Normal, Pearson Type III, and Log-Pearson Type III distributions. If at least one of the distributions was a good fit for each of the variables (R, I, and D), the threshold was considered optimal. Otherwise, the threshold was increased by 0.001, and the same steps were carried out until an optimal threshold was identified. All three variables (R, I, and D) were subjected to goodness-of-fit tests, as all are critical variables in event-based analysis. However, variable D was not included in the empirical copula

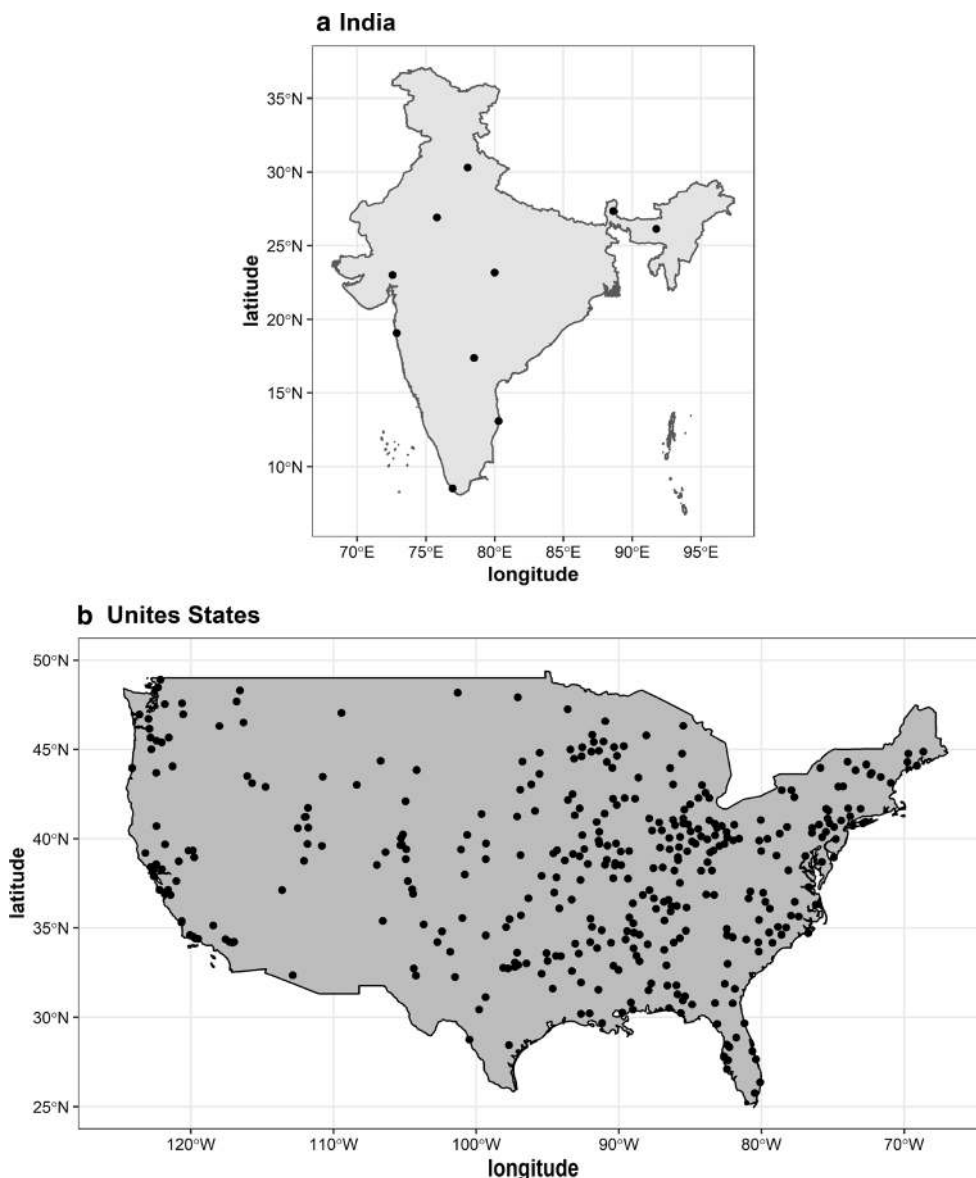


Fig. 1 Study locations in India and the United States used in this study

calculation, as  $D$  is mostly positively correlated to  $R$  and any rainfall events which possessed extreme nature in  $D$  also possessed extreme nature in  $R$ .

### 2.3 Derivation of dimensionless rainfall events

For a total depth ( $R$ ) and duration ( $D$ ) of any independent rainfall event, and a rainfall intensity  $r(d)$  at any instance of time,  $s$  during the rainfall, the accumulated rainfall  $R(d)$ , which is the RMC up to any duration  $d$  from the beginning of rainfall can be expressed as:

$$R(d) = \int_0^d r(s)ds \quad 0 \leq d \leq D \tag{1}$$

The variables  $R$  and  $D$  are random variables and vary in their magnitudes from event to event, making it challenging to develop a generalized model to represent the function  $R(d)$  of all the rainfall events. To eliminate the effect of the magnitude of random variables  $R$  and  $D$ , the RMC is used in its dimensionless form (Huff 1967; Bonta and Rao 1989):

$$M(t) = \int_0^t m(s)ds \quad 0 \leq t \leq 1, \quad 0 \leq M(t) \leq 1 \tag{2}$$

$$t = \frac{d}{D}, \quad M(t) = \frac{R(d)}{R(D)}, \quad m(s) = \frac{r(s)}{R(D)}$$

where  $M(t)$  is the dimensionless cumulative rainfall up to dimensionless time  $t$ , and  $m(s)$  is the dimensionless rainfall

at any instance. If  $n$  is the total number of time steps at which the event rainfall is recorded or available, then the intervals of dimensionless time,  $t$  ( $t_1, t_2, \dots, t_n$ ) will be at increments of  $1/n$  between 0 and 1. Then  $m(t)$  the rainfall occurring between any two-time steps can be written as

$$m(t_i) = M(t_i) - M(t_{i-1}) \quad i = 1, 2, \dots, n.$$

### 2.4 Lag-one correlation of RMC properties

The occurrence of rainfall in continuous time is stochastic and can be modeled as a Markov process (Sharma and Mehrotra 2010). The Markov property signifies that the probability of receiving any specific magnitude of rainfall in the current time step depends on the state of the rainfall in the previous time step. In this application, the DRMC is the process modeled. The property of the mass curve at any given point in time  $t_i$  has a dependence on its property on the preceding time step  $t_{i-1}$ . This study hypothesizes a linear dependence of properties of DRMC at any two consecutive time steps. A correlation analysis showed that using the combination, rainfall intensity ( $m(t_i)$ ) at any time step given the accumulated rainfall ( $M(t_{i-1})$ ) of the same rainfall event up to the previous time step is optimal for modeling the DRMC using a first-order Markov model. Few other combinations such as  $m(t_i)$  given  $m(t_{i-1})$ , a dimensionless process  $z(t)$  (Woolhiser and Osborn 1985) given  $z(t_{i-1})$ , and  $z(t_i)$  given  $M(t_{i-1})$  were also checked with linear correlation and examined for efficiency in modeling DRMC. However,  $m(t_i)$  given  $M(t_{i-1})$  was observed to be more appropriate.

### 2.5 First-order Markov model (FMM) for stochastic modeling of dimensionless rainfall mass curve

The methodology of stochastic modeling and generation of extreme rainfall events is described using a hypothetical DRMC (Fig. 2) as detailed in ensuing subsections.

### 2.5.1 Calculation of transition probability matrix

The DRMC was first divided into  $s$  equally spaced dimensionless rainfall states, such that the magnitude of dimensionless rainfall occurring between any two discrete time steps ( $m(t_i) = M(t_i) - M(t_{i-1})$ ) followed a uniform distribution. In other words, the number of bins of the histogram of rainfall pulses was adjusted so as to get a uniform distribution. A uniform distribution was used because the dimensionless rainfall pulses could be bounded between 0 and (1) These values then served as parameters of the uniform distribution, eliminating the need for parameter estimation. The use of any other distribution would require estimation of the parameters, further leading to parameter uncertainty. The resulting number of bins was then set as the number of dimensionless rainfall states ( $s$ ). From the analysis, a value of  $s = 11$  resulted in a uniform distribution of  $m(t_i)$  at most of the locations and was hence generalized to all study locations. Using a greater number of states would lead to a very large transition probability matrix, which is similar to overfitting the model and may also require a large quantity of data. A smaller number of states may not follow the uniform distribution of rainfall pulses. The states were: ten states at increments of  $m(t) = 0.1$  starting at 0.1; and, a special state (0) when  $M(t) = 0$  (Fig. 2). Once the exact number of states and the lower and upper bounds were identified, the dimensionless rainfall at each time step was converted into corresponding rainfall states in both DRMC and dimensionless hyetographs. For example (Fig. 2), at dimensionless time equals 0.5, the cumulative dimensionless rainfall was 0.61, corresponding to state 7. The amount of dimensionless rainfall occurring in the next time step (0.6) is 0.17, which corresponds to state 2, representing a transition between state 7 to state 2. The transition probabilities between all 11 dimensionless rainfall states were calculated and compiled as a transition probability matrix (TPM), as shown:

		TPM					Example Application							
		Future state (k)					Future state (k)							
		0	1	2	3	.	0	1	2	3	4			
$p =$	Current state (j)	0	$p_{0,0}$	$p_{0,1}$	$p_{0,2}$	$p_{0,3}$	.	0	0.1	0.3	0.3	0.2	0.1	=1
		1	$p_{1,0}$	$p_{1,1}$	$p_{1,2}$	$p_{1,3}$	.	1	0.2	0.1	0.2	0.3	0.2	=1
		2	$p_{2,0}$	$p_{2,1}$	$p_{2,3}$	$p_{2,3}$	.	= 2	0.1	0.4	0.3	0.1	0.1	=1
		3	$p_{3,0}$	$p_{3,1}$	$p_{3,2}$	$p_{3,3}$	.	3	0.2	0.2	0.3	0.1	0.2	=1
		.	.	.	.	.	.	4	0.1	0.3	0.2	0.3	0.1	=1

where  $p_{j,k} = \Pr\{m(t_i) = k | M(t_{i-1}) = j\}$ ,  $j, k = 1, 2, \dots, s$ , is the probability that an event receives a rainfall which falls in the state  $k$  at time step  $t_i$  given the actual occurred cumulative rainfall through time  $t_{i-1}$  falls in the state  $j$ . The matrix  $p$  satisfies the condition that  $p_{j,1} + p_{j,2} + \dots + p_{j,s} = 1$ , for all  $j$ . Further, a cumulative transition probability matrix (CTPM) was constructed from TPM, as shown:

		CTPM						Example Application					
		Future state ( $k$ )						Future state ( $k$ )					
		0	1	2	.	.	s	0	1	2	3	4	
Current state ( $j$ )	0	$P_{0,0}$	$P_{0,0}$	$P_{0,0}$	.	.	1	0	0.1	0.4	0.7	0.9	1
	1	$P_{1,0}$	$P_{1,0}$	$P_{1,0}$	.	.	1	1	0.2	0.3	0.5	0.8	1
	2	$P_{2,0}$	$P_{2,0}$	$P_{2,0}$	.	.	1	2	0.1	0.5	0.8	0.9	1
	3	$P_{3,0}$	$P_{3,0}$	$P_{3,0}$	.	.	1	3	0.2	0.4	0.7	0.8	1
	.	.	.	.	.	.	1	4	0.1	0.4	0.6	0.9	1

where  $P_{j,k} = \sum_{k=0}^s p_{j,k}$ . The CTPM was further used to generate stochastic DRMC

### 2.5.2 Stochastic generation of dimensionless rainfall mass curve

A dimensionless rainfall event was initiated with the current state  $j = 0$  at  $t = 0$  such that  $M(t = 0) = 0$ . A uniform distribution random number  $R_{u1} = U(0,1)$ , was then generated, and the future state  $k$  of dimensionless rainfall  $m(t_1)$  at the future time step  $t_1$  was identified such that  $P_{0,k-1} < R_{u1} \leq P_{0,k}$ . For example, considering the hypothetical CTPM, assuming  $R_{u1} = 0.21$ , then the state of the rainfall pulse at time step  $t_1$  will be, state = 1, given that state = 0 at  $t_0$ . This means that the magnitude of the dimensionless rainfall pulse that will occur at  $t_1$  will vary between 0 and 0.1 (bounds of state = 1). The width of each state, in this case, was 0.1; hence, the actual amount of rainfall  $m(t_1)$  at time step  $t_1$  was found by generating another uniform random number  $R_{u2} \sim U(0,0.1)$  and adding it to the lower bound (0 in this hypothetical case) of the generated state  $k$ . The cumulative rainfall  $M(t_1)$  was then found by adding  $m(t_1)$  to  $M(t_0) = 0$ , and the same steps were repeated until time step  $t_{n-1}$ . Further, the dimensionless rainfall  $m(t_n)$  of time step  $t_n$  was found as  $m(t_n) = 1 - M(t_{n-1})$ , to ensure the DRMC is bounded between 0 and 1. The generalized equation representing the stochastic generation of DRMC using the proposed method is given in Eq. 3

$$M(t_i) = M(t_{i-1}) + LB[P_{j,k}^{-1}(R_{u1} | M(t_{i-1}) = j)] + R_{u2} \tag{3}$$

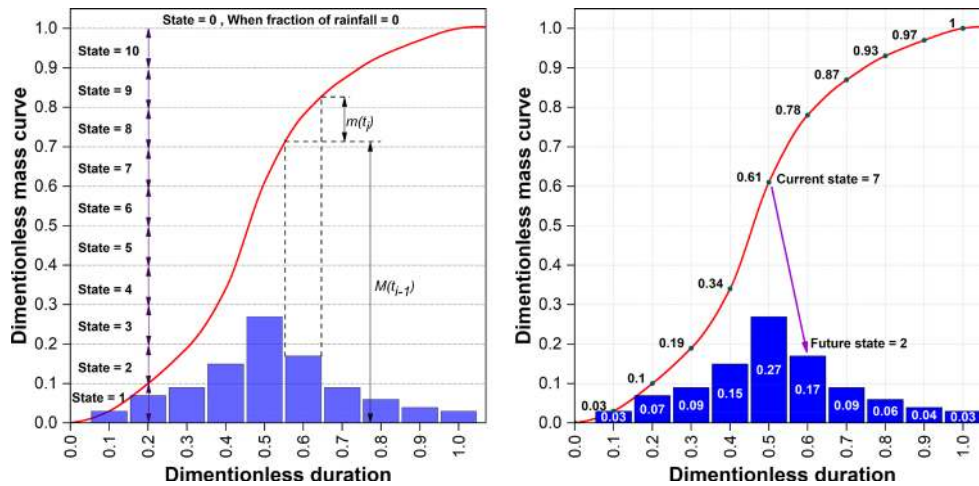
where LB represents the lower bound of the state of dimensionless rainfall.

## 2.6 Model performance evaluation

The performance of the model to characterize and generate DRMC was evaluated based on its ability to reproduce different characteristics of extreme rainfall events. Since the methodology in Sect. 2.5 is stochastic, a one to one comparison of observed and generated rainfall events may not be appropriate. Hence, at first, 100 synthetic DRMC were generated for each observed extreme rainfall event and converted into its non-dimensionless form by multiplying duration and total depth to the corresponding axis. Thus, the total number of extreme rainfall events becomes 100 times more than the observed extreme rainfall events. From these observed and simulated extremes, a quantile-quantile comparison of 1-h, 2-h, and 3-h peak rainfall intensities was carried out. Further, the total number of first, second, third, and fourth quartile storms and the dependence between peak rainfall intensity with total depth and duration of rainfall were also compared. Once the efficiency in reproducing rainfall properties had been examined, all the observed and simulated extreme rainfall events were run through a simple hydrologic model to estimate the direct runoff hydrograph and the return levels and return periods of peak discharges were compared.

## 2.7 Event-based hydrologic model

A simple event-based rainfall-runoff model using soil conservation service (SCS) method was implemented to generate hydrographs from the given rainfall mass curve



**Fig. 2** Hypothetical dimensionless rainfall mass curve (red curved line) and hyetographs (blue bars), with 11 dimensionless rainfall states. *Note:* The current state is always noted from the mass curve,

and the future state is noted from the hyetographs, thus, the current state could be lower or higher than the state in the next (future) time step

and catchment properties. In this rainfall-runoff model, curve number (CN) was used for estimating the rainfall excess and SCS dimensionless unit hydrograph to compute the direct runoff hydrograph at the catchment outlet. Since it is a comparative study, a simple hypothetical catchment with constant antecedent wetness was assumed for all the study locations. The assumed parameters of the catchment were as follows: time of concentration: 60 min; area of the catchment: 5 km<sup>2</sup>; CN: 95; initial abstractions: 10% of potential maximum retention; and, limb ratio of dimensionless unit hydrograph: 1.25. The value of limb ratio for urban areas was obtained from the Unit Hydrograph technical manual (National Operational Hydrologic Remote Sensing Center), and the time of concentration was assumed as 1 h since the minimum time resolution of available rainfall data for India is 1 h.

### 3 Results and discussion

#### 3.1 Preliminary assessments

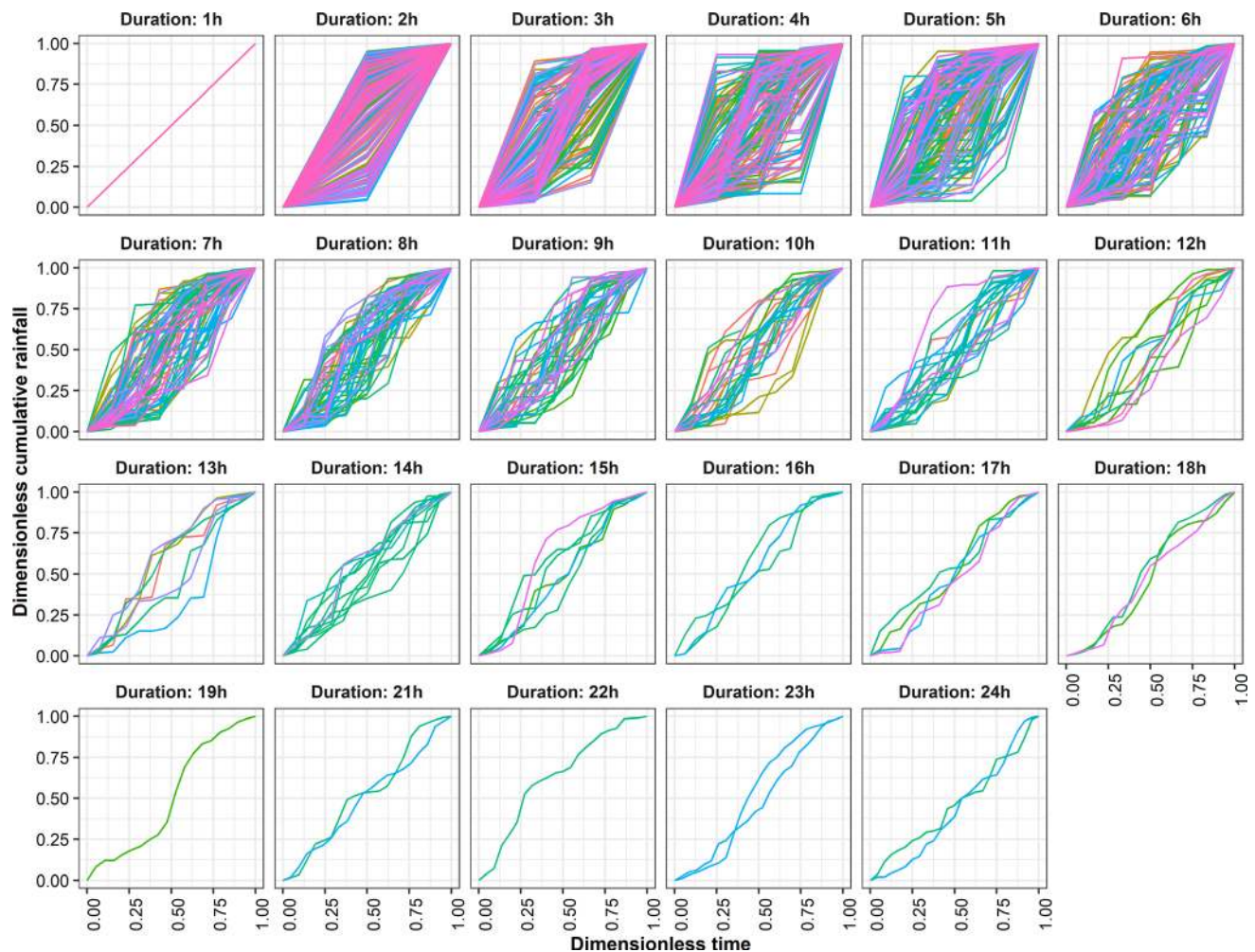
##### 3.1.1 Dependence of rainfall pattern with event duration

All the DRMC of different duration at Indian stations were plotted (Fig. 3) and visually examined for correlations between the duration of the extreme rainfall event and the pattern of the DRMC. It was observed that the extreme rainfall events of Indian cities could be classified into five groups based on the pattern of DRMC and the duration, which are: 1 h, 2 h, 3–6 h, 7–12 h, and 13 h and above. Since the data time resolution was 1 h, events of 1 h had uniform intensity, and the 2-h events had a unique pattern compared to the higher duration events with only one

intermediate time step. Extremes within the 3–6 h duration were evenly distributed, ranging from first to fourth quartile events. Extremes of 7–12 h had slightly less dominant peaks compared to 3–6 h events, as evidenced by relatively smoother DRMC (closer to 1:1 line) in Fig. 3, and extremes of greater than 13 h had more of uniform intensity in nature. Based on trial and error, and comparing performance, it was also noted that this duration classification was optimal for modeling DRMC. Hence, the DRMC of each of these duration categories were modeled and used separately for stochastically generating DRMC.

##### 3.1.2 Number of time intervals

A time step ( $n$ ) equal to the highest observed time step in each group was adopted for this study, i.e.,  $n = 2$  for 2 h events,  $n = 6$  for 3–6 h events  $n = 12$  for 7–12 h and  $n = 24$  for 13 h and above events. Time steps smaller than the time interval in the observed data require interpolations, which may lead to uncertainty, while longer time steps may not allow the shape of the mass curve to be preserved. A trial and error procedure confirmed that the selected values of  $n$  were optimal for all the groups and yielded better results than using a common number of time steps for all the events. A shape-preserving-piecewise-interpolation (Fritsch and Carlson 1980) was employed for all the interior events (3–5 h, 8–11 h, 13–23 h) to get the DRMC at the new time steps. The shape-preserving-piecewise-interpolation is a nonlinear interpolation method that avoids the assumption of uniform rainfall intensity within a time step and can preserve the monotonously increasing nature of DRMC.



**Fig. 3** Dimensionless rainfall mass curves at all the study locations of India. Different colors of the curves correspond to different stations. Note that the color scheme is used to help differentiate among the rainfall stations and does not influence the interpretation of the figure

### 3.1.3 Dimensionless rainfall events and Cumulative transition probability matrix (CTPM)

The dimensionless rainfall events (hyetographs) and the corresponding CTPM of the duration category 3–6 h at Gangtok are presented as heat plots (Fig. 4). Heat plots for all the remaining duration categories and stations are provided as supplementary materials (Figs. S1 and S2). The plotted extreme rainfall events (Fig. 4a) were sorted twice—first, in ascending order of time of peak rainfall intensity and second in descending order of magnitude of peak intensity within each peak intensity time step—before plotting them to achieve better inferential aesthetics. Almost 70 out of 127 extreme rainfall events (3–6 h events) received peak rainfall intensity between 0.3 and 0.7 of dimensionless time (Fig. 4a). The location of peak rainfall intensity is very important as it may affect the time and magnitude of the resulting peak runoff from the receiving catchment. More inference about the location of

the highest rainfall at different stations is given later in Sect. 3.2.2 (Fig. 5). In Fig. 4b, the color of the heat map indicates the cumulative transition probability corresponding to the combination of each rainfall state (current and future state). The current states of rainfall (Fig. 4b) ranged from 0 to 1 as it is always noted from the dimensionless mass curve. However, the next state or future state's highest value is always less than 1 as it is noted from the dimensionless hyetograph.

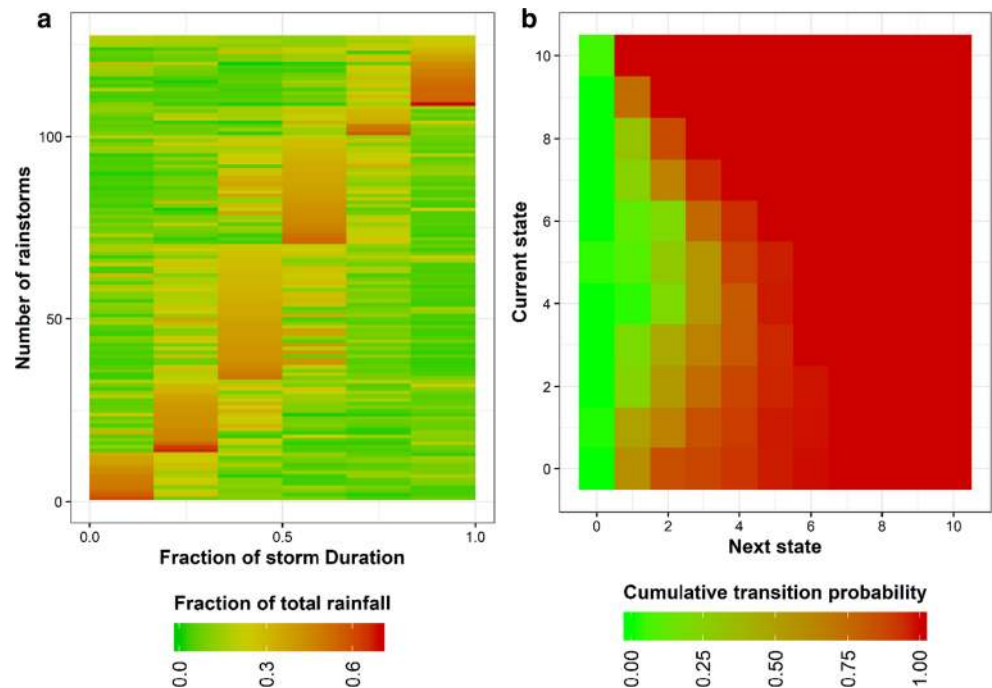
## 3.2 Detailed analysis

### 3.2.1 Comparison of 1, 2, 3-h peak rainfall intensity

Figure 6 presents comparisons of peak rainfall intensities for all the stations along with the one of the widely used performance measures: Nash-Sutcliffe-Efficiency (NSE). The value of NSE was calculated between 100 quantiles of both observed, and the stochastically generated data. The



**Fig. 4** **a** Dimensionless rainfall events, and **b** corresponding cumulative transition probability matrix for the rainfall events of category 3–6 h at Gangtok station



average of NSE across all the stations was: 0.87 (1 h); 0.93 (2 h); and 0.95 (3 h). The increase in average NSE with an increase in the duration of peak rainfall intensity shows evidence of better reproducibility of higher duration intensities than that of lower duration intensities. Nevertheless, all three duration peak intensities are well replicated by the model with the NSE value of at least 0.8 in all the cases. Further, the consistency of performance across all stations can be inferred based on the standard deviations (SD) of the NSE (0.04 (1 h); 0.02 (2 h); and, 0.02 (3 h)) which were low.

### 3.2.2 Comparison of fraction of quartile storms

Figure 5 presents a comparison of the observed and generated fractions of the total number of storms in each quartile. The green and red dots in Fig. 5 indicate the observed and simulated fraction of quartile storms, respectively. The gap between green and red dots gives the difference in the fraction of storms; for an efficient model, this gap should be as small as possible. A red dot to the left of the green indicates that the simulated number of that specific quartile storms is less than the observed and vice versa. The mean error in the representation of first, second, third, and fourth quartile was 2%, 1%, 2%, and – 5%, respectively. The percentage of storms that were underestimated (+) or overestimated (–) was low in all the four quartiles indicating that the proposed method was able to reproduce the number of storms in each quartile fairly well.

From Fig. 5, from among 10 stations, at 9 stations, more than 25% of the total storms were of the first quartile. The percentage was more than 30% at 6 stations and more than 40% at 2 stations. On average, more than 60% of the extreme rainfall events had heavy rainfall in the first half of the storm. However, at Jaipur and Hyderabad, more than 45%, and at Chennai, more than 35% of storms were first quartile storms, which means most of the rainfall occurred at the beginning of the storm. The first two stations fall under arid climates with mostly convective rainfall patterns, and the latter is a coastal city where cyclonic effects also play a significant role. At Gangtok and Bombay, most of the extremes were of second and third quartile storms indicating the highest rainfall occurred during the middle of the events. The south-west monsoon is a major rainfall contributor at these two stations. Gangtok falls under the mountainous region, and Bombay is a coastal city. Based on the analysis, the association between local climate and the pattern of rainfall at most of the stations does not appear to be distinctive. Detailed analysis with data from a large number of stations may be required to arrive at a definitive conclusion.

### 3.2.3 Correlation of peak rainfall intensity with duration and total depth of rainfall

Table 1 presents the results of the Kendall correlation (K) for both combinations, along with their errors, mean, and standard deviation of error across all the stations. The results of Table 1 indicate that the correlation of I and R,

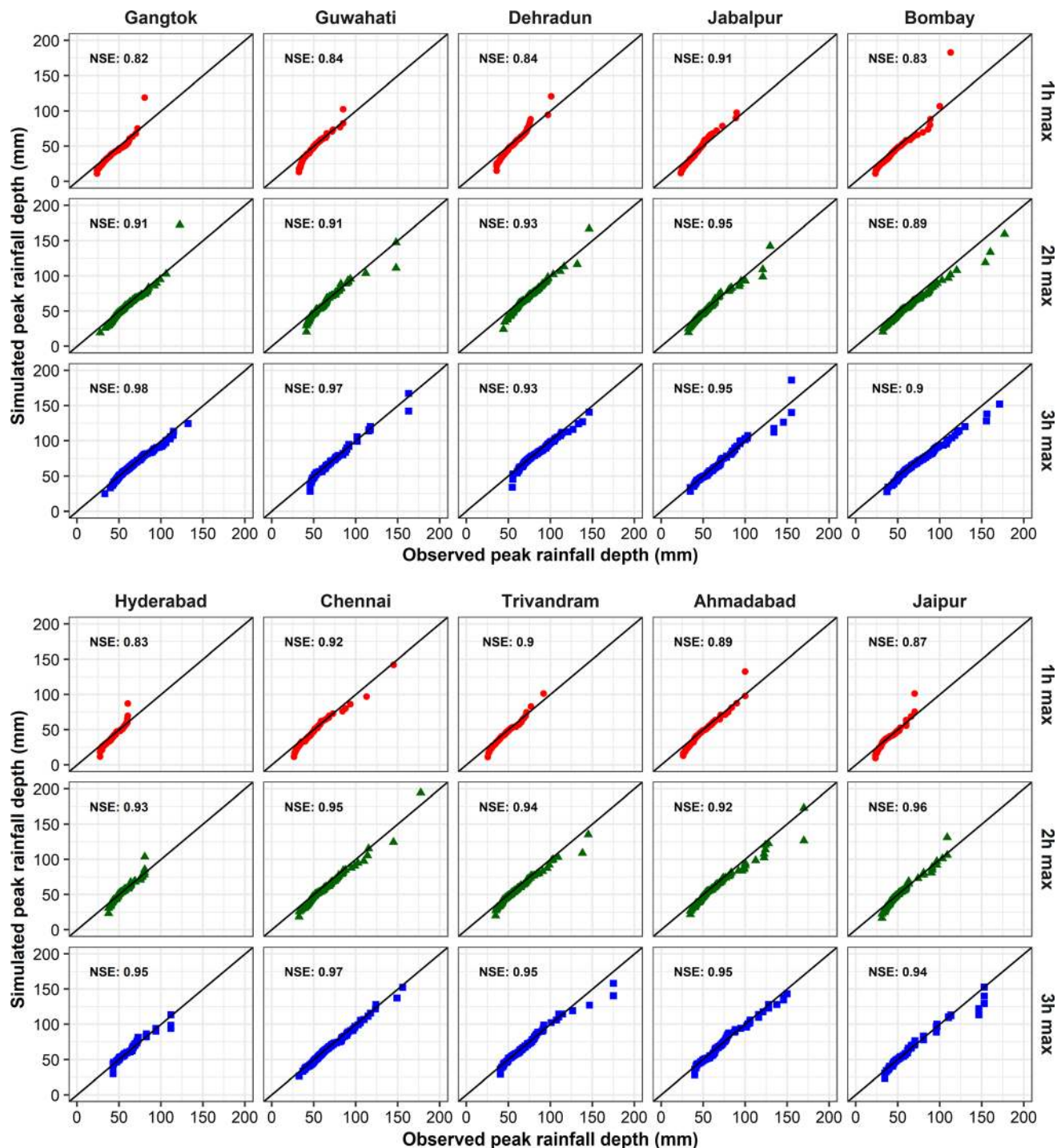


Fig. 5 Dumbbell plots showing the difference between observed and the simulated fraction of storms in each quartile

and I and D of rainfall events were replicated fairly well and consistent at different stations.

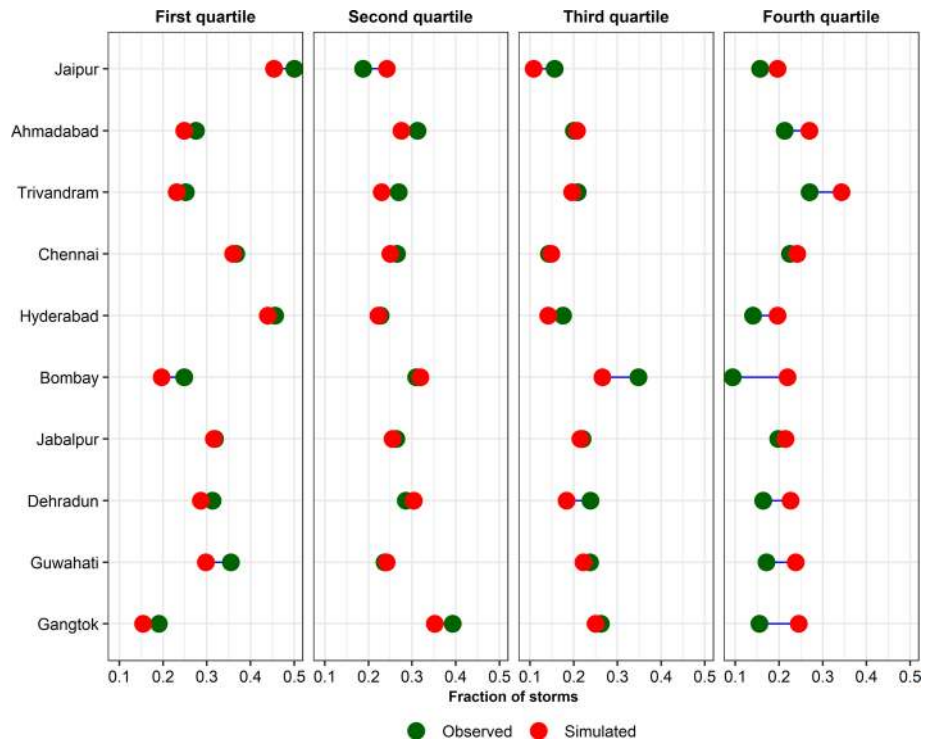
Based on comparisons of: 1, 2, and 3-h maximum rainfall depth; number of first, second, third and fourth quartiles events; and the correlations of I with D and R, the developed FMM for the stochastic generation of time distribution of rainfall events could replicate the observed

time distributions at all the 10 study stations with sufficient accuracy.

### 3.2.4 Peak discharge and its return period

All the observed extreme rainfall events and the stochastically generated extreme rainfall events were run through

**Fig. 6** Scatter plot of peak rainfall intensities obtained from observed and simulated extreme rainfall events for 1-h, 2-h, and 3-h duration at study locations in India



**Table 1** Error in the correlation between event duration and peak intensity, and total rainfall depth and peak intensity of observed and stochastically generated rainfall

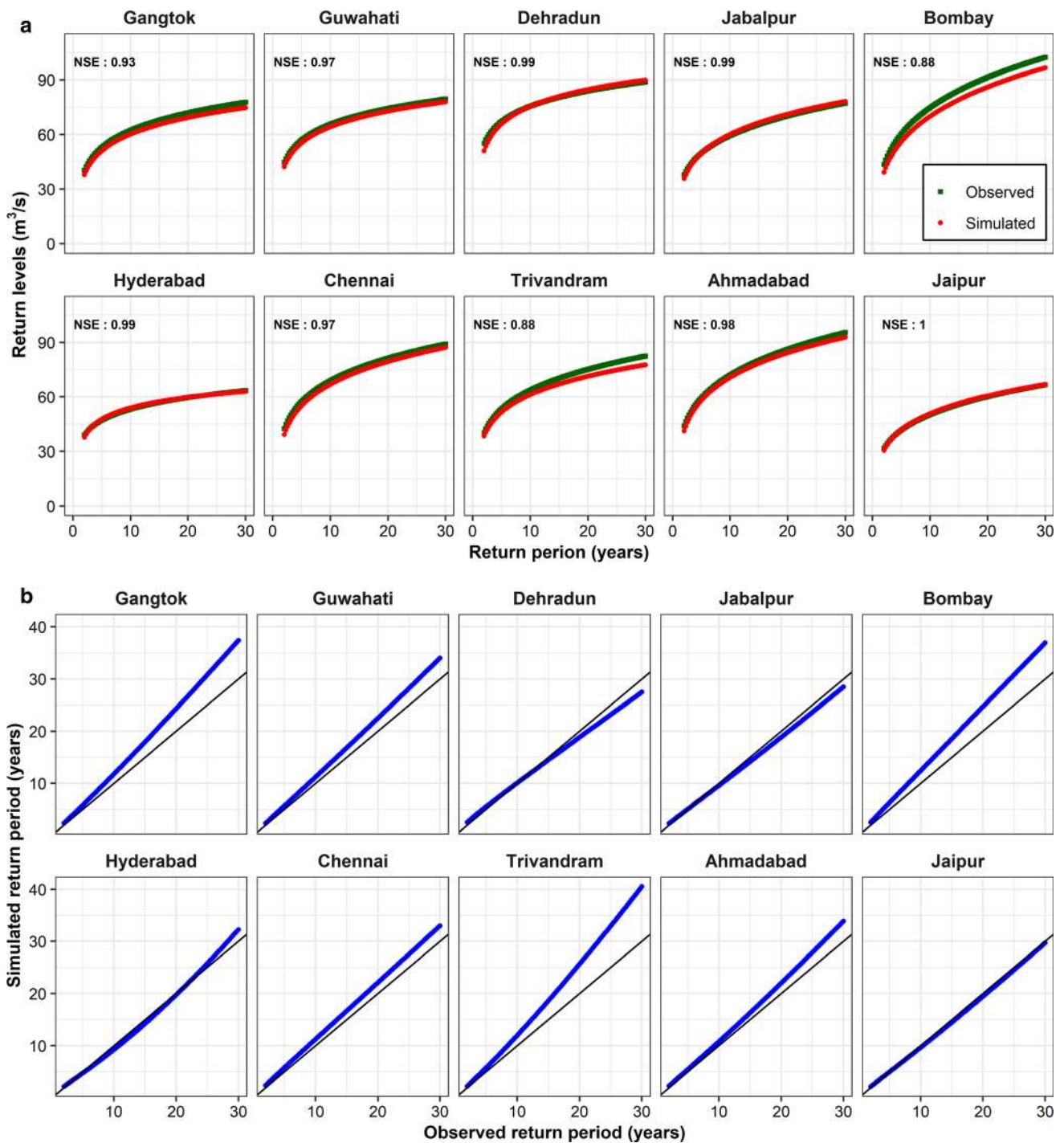
City	Duration vs. peak intensity			Total depth vs. peak intensity		
	$K_{obs}$	$K_{sim}$	$E = K_{obs} - K_{sim}$	$K_{obs}$	$K_{sim}$	$E = K_{obs} - K_{sim}$
Gangtok	0.23	0.32	- 0.09	- 0.17	- 0.17	0.00
Guwahati	0.26	0.25	0.01	- 0.18	- 0.34	0.16
Dehradun	0.01	0.11	- 0.10	- 0.23	- 0.28	0.05
Jabalpur	0.14	0.17	- 0.03	- 0.22	- 0.24	0.02
Bombay	0.20	0.22	- 0.02	- 0.08	- 0.12	0.04
Hyderabad	0.16	0.19	- 0.03	- 0.16	- 0.27	0.11
Chennai	0.17	0.17	0.00	- 0.22	- 0.24	0.02
Trivandrum	0.13	0.11	0.02	- 0.21	- 0.29	0.08
Ahmadabad	0.14	0.28	- 0.14	- 0.21	- 0.13	- 0.08
Jaipur	0.15	0.18	- 0.03	- 0.19	- 0.28	0.09
Mean			- 0.04			0.05
SD			0.05			0.07

*K* Kendall correlation

an event-based hydrological model (Sect. 2.7) to obtain the resulting direct runoff hydrograph. The magnitude and return periods of peak runoff from both observed and simulated rainfall events were further compared for accuracy in replication. Two types of peak runoff comparisons were carried out: (1) peak runoff of different return periods; and, (2) difference in return periods of discharge obtained from observed and synthetic rainfall extremes. Since the design of most of the structures related to urban stormwater management adopts return periods between 2 and 30 years, the comparison was carried out for 2–30-year

return periods. For the estimation of return level and the return period, a generalized extreme value distribution was observed to fit well for peak discharge at all the study locations.

Figure 7a presents the results of the comparison of peak runoff and its return periods for 2–30-year return periods. From Fig. 7a, though the peak discharges from simulated rainfall was lower than the peak discharges from the observed rainfall at some stations, the differences in magnitude was low. The mean and SD of the difference in peak runoff calculated across all the stations was 1.56 and



**Fig. 7** Comparison of return levels and return periods of peak discharge from observed and simulated rainfall events

$2.11 m^3/s$ , respectively. The mean difference observed at Bombay ( $4.9 m^3/s$ ) was the highest, followed by Trivandram ( $4.4 m^3/s$ ). Jabalpur had the lowest mean difference ( $-0.6 m^3/s$ ). Further, a comparison of concurrent return periods of peak discharges from observed and simulated rainfall events was carried out to check the changes in frequencies of peak discharge. These return period and

return period comparisons are presented in Fig. 7b. The increase in return periods of peak runoff from simulated events compared to peak runoff from observed events was due to the underestimation of peak discharge (Fig. 7a). Though the difference in return periods seems high, the difference in corresponding peak runoff was low. For instance, a 30-year observed event at Bombay had a 36.7-

year return period simulated, but the difference in peak discharge was only 5.84 m<sup>3</sup>/s (5.7%, from 102.42 to 96.58 m<sup>3</sup>/s) and, similarly, 6.58 m<sup>3</sup>/s (8%, from 82.53 to 75.95 m<sup>3</sup>/s) for a discrepancy in return period from 30 years (observed) to 46.23 years (simulated) at Trivandram. However, overall, the properties of both rainfall and peak discharges were well replicated by the model.

### 3.3 Application to 15 min time interval extremes in the United States

Extreme rainfall events at US stations were identified in a manner similar to that for India stations, except that the peak rainfall intensities were taken for 15-mins. It was noted that the pattern of dimensionless rainfall mass curves in the US stations had a lower dependence on the duration of the rainfall. Hence all the extreme events at each station were modeled together irrespective of the duration. The number of time steps equal to 12 was observed to perform better and the number of rainfall states used was 11. The results of performance are presented in Figs. 8 and 9 as the mean and standard deviation of NSE for different stations within each climate region of the US. The classification of climate regions was obtained from Karl and Koss (1984). From Fig. 8, the mean of NSE of peak rainfall intensity and peak discharge at different climate regions varied from 0.65 to 0.99, and the standard deviation from 0 to 0.25. Similar to India stations, fourth quartile storms were slightly overestimated at all the locations (Fig. 9). However, the fraction of overestimation was less than 0.2 in most of the cases. Lower values of standard deviation (Figs. 8 and 9) indicate that the performance of the model was consistent at all the study locations.

## 4 Discussion

The time distribution of extreme rainfall events is one of the preeminent properties that govern the design of various urban stormwater management structures. A model used for characterizing these properties must be robust and computationally efficient while requiring a small number of parameters to produce satisfactory results. Model development should also entail a comprehensive performance evaluation of the developed methods to characterize the time distribution of extreme rainfall events. In this study, utilizing the inherently stochastic nature of rainfall occurrence, a first-order multistate Markov model was developed for modeling the time distribution of dimensionless extreme rainfall events. The developed model was applied to hourly rainfall data at ten stations in India and tested on 15 min rainfall data at 205 stations in the United States. Further, a comprehensive performance evaluation was carried out to verify the efficiency of the model. The performance comparisons for India (US) include: 1-h (15-min), 2-h (30-min), 3-h (45-min) peak rainfall intensities; fraction of first, second, third, and fourth quartile storms; the correlation of peak rainfall intensity with duration and total rainfall depth; and, return level and return periods of peak discharges when the simulated extreme rainfall events were applied over a hypothetical urban catchment. Results indicated that the model was able to reproduce the time behavior of extreme rainfall events fairly well.

Identification of extreme rainfall events mainly requires a threshold to define a rainfall pulse as rainy or non-rainy, an event separation time to delineate independent rainfall events and, extraction of extremes from among all the rainfall events. Based on (Subramanya 1994; Deodhar 2008), precipitation of intensity of at least 1 mm/h was classified as rainfall, below which any precipitation was

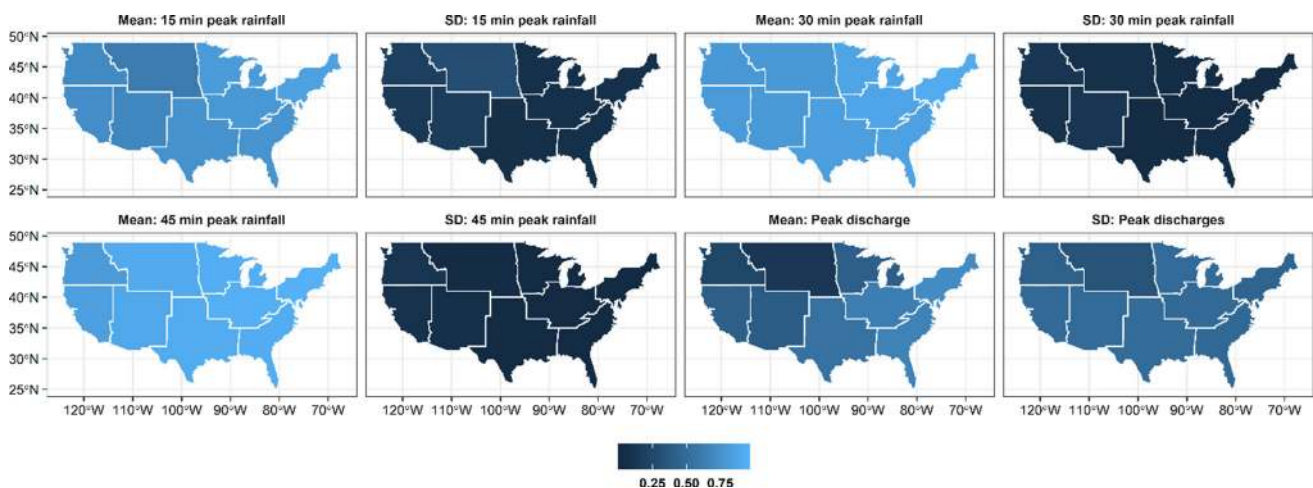
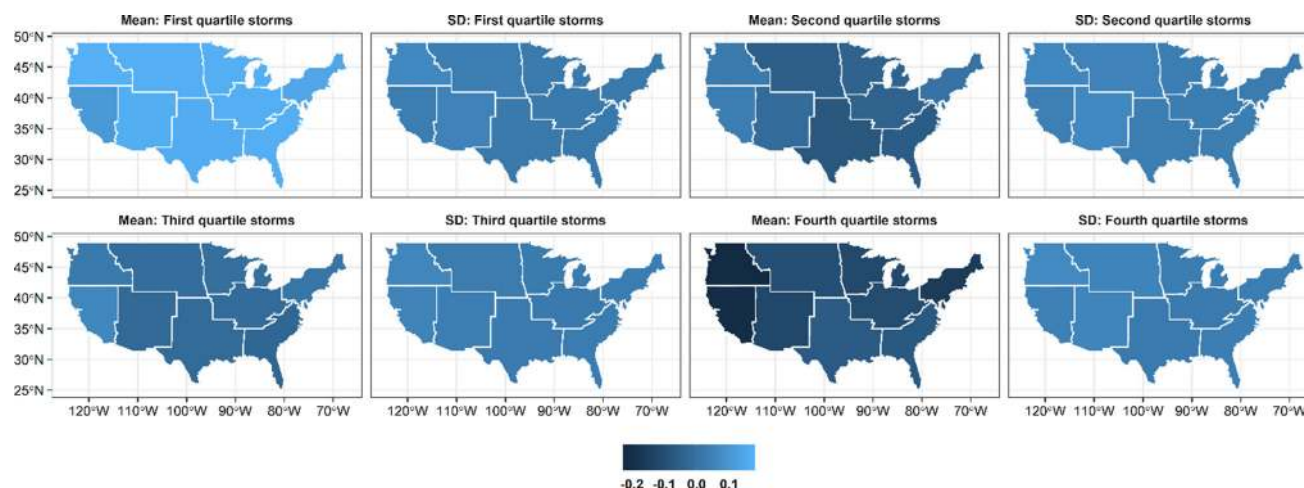


Fig. 8 Mean and standard deviation of NSE of all the US stations in each climate region



**Fig. 9** Mean and standard deviation of the error in fraction of first to fourth quartile storms at all the US stations in each climate region

considered a drizzle (Subramanya 1994; Deodhar 2008). Selecting a different threshold would generally not affect the study of extreme rainfall behavior, as extremes possess a higher magnitude and intensity of rainfall. Event separation time ranging from 1 to 12 h, is common in urban-based studies (Park et al. 2013; Balistrocchi and Bacchi 2017; Jun et al. 2017; Nojumuddin et al. 2018). However, a minimum of 3 h (Adams and Papa 2000) and a maximum of 12 h (Hassini and Guo 2016) are recommended to avoid the separation of correlated rainfall events and the grouping of events that are too far apart, respectively. There are many methods developed to identify the appropriate event separation time based on the literature (Joo et al. 2013; Lee and Kim 2018). However, the commonly-used 6-h separation time (Kao and Govindaraju 2007; Bezak et al. 2016) was adopted in this study.

Further, to extract only extreme rainfall events, univariate annual maximum or peak-over-threshold (Coles 2001) methods that consider only one variable (usually peak intensity) are generally used. However, an approach that considers multiple properties of rainfall would be more appropriate to define the extreme nature of rainfall, because, based on the intended design problem, different properties of extreme rainfall events (R, I and, D) will govern the design (Kao and Govindaraju 2007). In the design of urban water management structures, the total magnitude of rainfall and the peak intensity may have a higher impact on the design. Hence a bivariate peak-over-threshold with total rainfall and peak intensity as critical variables was adopted in this study to extract extreme rainfall events.

An efficient stochastic RMC generation model must preserve the joint behavior of the different properties of a rainfall event. In this study, a comparison of linear correlation was carried out between I and R, and I and D of observed and simulated rainfall events. A Kendall's

correlation was selected as an index to measure the correlation as it is a non-parametric measure and does not assume normality of the error. However, the use of other indexes such as Pearson and Spearman correlation could also result in similar comparisons. The error in the Kendall correlation of observed and simulated extreme events properties was less than 0.1 at more than 80% of the stations.

The quartile of an RMC gives an idea of the shape of the mass curve and the quarter of the event duration where the maximum amount of rainfall occurred. First quartile storms usually receive most of their rainfall in the first half of the storm resulting in an early peak in the DRH, while fourth quartile storms receive most of their rain in the second half, causing a delayed peak. Conversely, fourth quartile storms might generate higher peaks than first quartile storms as the catchment response increases with increasing duration of rainfall due to minimized losses. It is, thus, very important that generated synthetic RMCs accurately reproduce the number of storms in different quartiles. In our study, the Markov model tended to overestimate the number of fourth quartile storms at most stations. This overestimation may lead to a higher number of delayed peak discharges, and higher magnitude peak discharges being simulated compared with those obtained from observed rainfall events.

When both observed and simulated rainfall events were run through a simple hypothetical urban catchment, the resulting discharges corresponded well. However, at a few stations (for example, Trivandrum) a small difference ( $7 \text{ m}^3/\text{s}$ ) in discharge led to a higher difference in return period (16 yrs). The higher difference in the return period for a small difference in peak discharge was possibly due to the small value of the GEV shape parameter leading to a thin and asymptotic upper tail, or due to a negative shape parameter in simulated peak discharge, while the observed had a positive one. The shape of the upper tail of the

distribution of peak discharge could be attributed to the simulation method for the rainfall hyetograph, from which the overestimation of the number of fourth quartile storms (Fig. 5) resulted in a higher number of delayed and higher magnitude peak discharges. This overestimation of fourth quartile storms and the differences in the peak discharges are mainly due to differences in the time distribution of rainfall as all other variables are kept the same for both observed and simulated events. Though the simulated rainfall events led to higher peak discharges than did the observed rainfall events, the mean of the peak discharge was lower in the discharges from simulated events at all the stations. This reduced mean is the main reason for the reduced return levels of discharges from simulated rainfall events.

Generally, the properties of rainfall events are modeled to identify the resulting peak discharge, which is a deciding variable for the design of various urban water management structures. From a comprehensive performance evaluation, the multistate FMM was able to reproduce most of the properties of extreme rainfall events and thereby to increase the quality of the input rainfall hyetograph to hydrological models. The proposed model efficiently captures any possible shape of observed dimensionless extreme rainfall events, as stochastic occurrences of rainfall are generated at each time step. In this study, empirically calculated transition probabilities rather than estimated parameters were used, which eliminated uncertainty due to parameter estimation. Generally, the quantity and the quality of the data used, and the presence of sudden bursts in the hyetographs (as might be the case with outliers or rare events) introduces uncertainty in Markov chain-based models. To come to a definitive conclusion, a separate uncertainty analysis would be needed. Such an analysis is, however, beyond the scope of the current study.

## 5 Conclusions

A stochastic model based on first-order multistate Markov model was developed to characterize the time distribution of a dimensionless form of extreme rainfall events. The methodology was tested on hourly rainfall data at 10 stations in India, and 15-min rainfall data at 205 stations in the United States. The model was able to reproduce peak rainfall intensities at several locations in India and the US with high efficiency. The percentage of rainfall events falling in first to fourth quartile storms were also simulated with a low error at almost all the locations. A comparison of the correlation of peak rainfall intensity with total depth and duration of rainfall also showed the error in the correlation was minimal. In general, the proposed method was able to characterize the time behavior of extreme rainfall

events very well at a wide range of geographical and climatological regions, which is important for designing various urban water resources management structures. The return levels and return periods of the peak discharges obtained using simulated extreme rainfall events matched well with those obtained using observed data. Accurate estimation of peak discharges can reduce problems related to under design, such as flooding and those related to over design, which may lead to uneconomical construction and maintenance. The results of this study are specific to India, and the US. Methodologies and approaches developed are widely applicable.

**Acknowledgements** The 15 min rainfall data for the US stations that support the findings of this study are openly available from the National Oceanic and Atmospheric Administration (NOAA), at <https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ncdc:C00505#>, dataset identifier [gov.noaa.ncdc:C00505].

**Author contributions** All authors contributed to the study conceptualization and design. Material preparation, data collection, analysis, and writing of the first draft of the manuscript were performed by RAN. Data collection, funding, supervision, and reviewing and editing of the manuscript were performed by MWG, IC, and KPS. All authors have read and approved the final manuscript.

**Funding** This work was funded by the Science and Engineering Research Board, Department of Science and Technology, Government of India, and Purdue University, United States, through the SERB-Purdue Overseas Visiting Doctoral Fellowship and in part by USDA National Institute of Food and Agriculture, Hatch Project IND00000752.

**Availability of data and materials** The hourly rainfall data for Indian stations that support the findings of this study are available from India Meteorological Department (IMD) Pune, Ministry of Earth Science, Government of India. Restrictions apply to the availability of these data. The data were used under license for this study. Data are available at IMD Data Supply Portal with the permission of IMD Pune.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Consent to participate/for publication** All authors have consented to participate and the publication of the manuscript.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright

holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Acreman MC (1990) A simple stochastic model of hourly rainfall for Farnborough, England. *Hydrol Sci J* 35:119–148. <https://doi.org/10.1080/02626669009492414>
- Adams BJ, Papa F (2000) *Urban Stormwater management planning with analytical probabilistic models*. Wiley, New York
- Awadallah AG, Younan NS (2012) Conservative design rainfall distribution for application in arid regions with sparse data. *J Arid Environ* 79:66–75. <https://doi.org/10.1016/j.jaridenv.2011.11.032>
- Azarnivand A, Camporese M, Alaghmand S, Daly E (2020) Simulated response of an intermittent stream to rainfall frequency patterns. *Hydrol Process* 34:615–632. <https://doi.org/10.1002/hyp.13610>
- Balistrocchi M, Bacchi B (2017) Derivation of flood frequency curves through a bivariate rainfall distribution based on copula functions: application to an urban catchment in northern Italy's climate. *Hydrol Res* 48:749–762. <https://doi.org/10.2166/nh.2017.109>
- Beck HE, Zimmermann NE, McVicar TR et al (2018) Present and future Köppen–Geiger climate classification maps at 1-km resolution. *Sci Data* 5:180214. <https://doi.org/10.1038/sdata.2018.214>
- Berk M, Špačková O, Straub D (2017) Probabilistic design storm method for improved flood estimation in ungauged catchments. *Water Resour Res* 53:10701–10722. <https://doi.org/10.1002/2017WR020947>
- Bezak N, Šraj M, Mikoš M (2016) Copula-based IDF curves and empirical rainfall thresholds for flash floods and rainfall-induced landslides. *J Hydrol* 541:272–284. <https://doi.org/10.1016/j.jhydrol.2016.02.058>
- Bonta JV (2004) Development and utility of Huff curves for disaggregating precipitation amounts. *Appl Eng Agric* 20:641–653. <https://doi.org/10.13031/2013.17467>
- Bonta JV, Rao AR (1989) Regionalization of storm hyetographs. *JAWRA J Am Water Resour Assoc* 25:211–217. <https://doi.org/10.1111/j.1752-1688.1989.tb05683.x>
- Coles S (2001) *An introduction to statistical modeling of extreme values*. Springer, London
- Deodhar MJ (2008) *Elementary engineering hydrology*. Pearson Education, Pearson
- Di Baldassarre G, Brath A, Montanari A (2006) Reliability of different depth-duration-frequency equations for estimating short-duration design storms. *Water Resour Res* 42:1–6. <https://doi.org/10.1029/2006WR004911>
- Dullo TT, Kalyanapu AJ, Teegavarapu RSV (2017) Evaluation of changing characteristics of temporal rainfall distribution within 24-hour duration storms and their influences on peak discharges: case study of Asheville, North Carolina. *J Hydrol Eng* 22(11):05017022
- Entekhabi D, Rodriguez-Iturbe I, Eagleson PS (1989) Probabilistic representation of the temporal rainfall process by a modified Neyman–Scott rectangular pulses model: parameter estimation and validation. *Water Resour Res* 25:295–302. <https://doi.org/10.1029/WR025i002p00295>
- Fritsch FN, Carlson RE (1980) Monotone piecewise cubic interpolation. *SIAM J Numer Anal* 17:238–246. <https://doi.org/10.1137/0717021>
- Garcia-Guzman A, Aranda-Oliver E (1993) A stochastic model of dimensionless hyetograph. *Water Resour Res* 29:2363–2370. <https://doi.org/10.1029/93WR00517>
- Ghassabi Z, Kamali GA, Meshkati AH et al (2016) Time distribution of heavy rainfall events in south west of Iran. *J Atmos Solar Terrest Phys* 145:53–60. <https://doi.org/10.1016/j.jastp.2016.03.006>
- Guo A, Chang J, Wang Y et al (2018) Bivariate frequency analysis of flood and extreme precipitation under changing environment: case study in catchments of the Loess Plateau, China. *Stoch Environ Res Risk Assess* 32:2057–2074. <https://doi.org/10.1007/s00477-017-1478-9>
- Guo JC, Hargadin K (2009) Conservative design rainfall distribution. *J Hydrol Eng* 14:528–530. [https://doi.org/10.1061/\(asce\)he.1943-5584.0000013](https://doi.org/10.1061/(asce)he.1943-5584.0000013)
- Gupta VK, Waymire EC, Gupta VK, Waymire EC (1993) A statistical analysis of mesoscale rainfall as a Random Cascade. *J Appl Meteorol* 32:251–267. [https://doi.org/10.1175/1520-0450\(1993\)032<0251:ASAOMR>2.0.CO;2](https://doi.org/10.1175/1520-0450(1993)032<0251:ASAOMR>2.0.CO;2)
- Hassini S, Guo Y (2016) Exponentiality test procedures for large samples of rainfall event characteristics. *J Hydrol Eng* 21:1–12. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001352](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001352)
- Hingray B, Monbaron E, Jarrar I et al (2002) Stochastic generation and disaggregation of hourly rainfall series for continuous hydrological modelling and flood control reservoir design. *Water Sci Technol* 45:113–119. <https://doi.org/10.2166/wst.2002.0035>
- Huff FA (1967) Time distribution of rainfall in heavy storms. *Water Resour Res* 3:1007–1019. <https://doi.org/10.1029/WR003i004p01007>
- Huff FA (1990) Time distribution of heavy rainstorms in Illinois. *Illinois state water survey, Circular 173*, 1990
- Joo J, Lee J, Kim J et al (2013) Inter-event time definition setting procedure for urban drainage systems. *Water* 6:45–58. <https://doi.org/10.3390/w6010045>
- Jun C, Qin X, Gan TY et al (2017) Bivariate frequency analysis of rainfall intensity and duration for urban stormwater infrastructure design. *J Hydrol* 553:374–383. <https://doi.org/10.1016/j.jhydrol.2017.08.004>
- Jun C, Qin X, Tung Y-K, De Michele C (2018) Storm event-based frequency analysis method. *Hydrol Res* 49:700–710. <https://doi.org/10.2166/nh.2017.175>
- Kao SC, Govindaraju RS (2007) A bivariate frequency analysis of extreme rainfall with implications for design. *J Geophys Res Atmos* 112:1–15. <https://doi.org/10.1029/2007JD008522>
- Kao SC, Govindaraju RS (2008) Trivariate statistical analysis of extreme rainfall events via the Plackett family of copulas. *Water Resour Res* 44:1–19. <https://doi.org/10.1029/2007WR006261>
- Karl TR, Koss WJ (1984) *Regional and national monthly, seasonal, and annual temperature weighted by area, 1895–1983. Historical Climatology Series 4–3*, National Climatic Data Center, Asheville
- Kimoto A, Canfield HE, Stewart D (2011) Comparison of synthetic design storms with observed storms in Southern Arizona. *J Hydrol Eng* 16:935–941. [https://doi.org/10.1061/\(asce\)he.1943-5584.0000390](https://doi.org/10.1061/(asce)he.1943-5584.0000390)
- Kossieris P, Makropoulos C, Onof C, Koutsoyiannis D (2018) A rainfall disaggregation scheme for sub-hourly time scales: coupling a Bartlett–Lewis based model with adjusting procedures. *J Hydrol* 556:980–992. <https://doi.org/10.1016/j.jhydrol.2016.07.015>
- Kottegoda NT, Natale L, Raiteri E (2003) A parsimonious approach to stochastic multisite modelling and disaggregation of daily rainfall. *J Hydrol* 274:47–61. [https://doi.org/10.1016/S0022-1694\(02\)00356-6](https://doi.org/10.1016/S0022-1694(02)00356-6)



- Koutsoyiannis D, Mamassis N (2001) On the representation of hyetograph characteristics by stochastic rainfall models. *J Hydrol* 251:65–87. [https://doi.org/10.1016/S0022-1694\(01\)00441-3](https://doi.org/10.1016/S0022-1694(01)00441-3)
- Lee EH, Kim JH (2018) Development of new inter-event time definition technique in Urban Areas. *KSCE J Civ Eng* 22:3764–3771. <https://doi.org/10.1007/s12205-018-1120-5>
- Marani M (2003) On the correlation structure of continuous and discrete point rainfall. *Water Resour Res*. <https://doi.org/10.1029/2002WR001456>
- Molnar P, Burlando P (2005) Preservation of rainfall properties in stochastic disaggregation by a simple random cascade model. *Atmos Res* 77:137–151. <https://doi.org/10.1016/j.atmosres.2004.10.024>
- National Operational Hydrologic Remote Sensing Center NWS Unit Hydrograph Technical Manual. [https://www.nohrsc.noaa.gov/technology/gis/uhg\\_manual.html](https://www.nohrsc.noaa.gov/technology/gis/uhg_manual.html). Accessed 26 Aug 2019
- Nelsen RB (2006) An introduction to copulas. Springer, New York
- Nguyen V-T-V, Rousselle J (1981) A stochastic model for the time distribution of hourly rainfall depth. *Water Resour Res* 17:399–409. <https://doi.org/10.1029/WR017i002p00399>
- Nojumuddin NS, Yusof F, Yusop Z (2018) Determination of minimum inter-event time for storm characterisation in Johor, Malaysia. *J Flood Risk Manag* 11:S687–S699. <https://doi.org/10.1111/jfr3.12242>
- Olsson J, Burlando P (2002) Reproduction of temporal scaling by a rectangular pulses rainfall model. *Hydrol Process* 16:611–630. <https://doi.org/10.1002/hyp.307>
- Onof C, Chandler RE, Kakou A et al (2000) Rainfall modelling using Poisson-cluster processes: a review of developments. *Stoch Environ Res Risk Assess* 14:0384–0411. <https://doi.org/10.1007/s004770000043>
- Onof C, Wheater HS (1994) Improvements to the modelling of British rainfall using a modified Random Parameter Bartlett–Lewis Rectangular Pulse Model. *J Hydrol* 157:177–195. [https://doi.org/10.1016/0022-1694\(94\)90104-X](https://doi.org/10.1016/0022-1694(94)90104-X)
- Park M, Yoo C, Kim H, Jun C (2013) Bivariate frequency analysis of annual maximum rainfall event series in Seoul, Korea. *J Hydrol Eng* 19:1080–1088. [https://doi.org/10.1061/\(asce\)he.1943-5584.0000891](https://doi.org/10.1061/(asce)he.1943-5584.0000891)
- Rodríguez-Iturbe I, de Power BF, Valdés JB (1987) Rectangular pulses point process models for rainfall: analysis of empirical data. *J Geophys Res* 92:9645. <https://doi.org/10.1029/JD092iD08p09645>
- Rodriguez-Iturbe I, Cox DR, Isham V (1988) A point process model for rainfall: further developments. *Proc R Soc Lond A Math Phys Sci* 417:283–298. <https://doi.org/10.1098/rspa.1988.0061>
- Sampson AA, Wright DB, Stewart RD, LoBue A (2020) The role of rainfall temporal and spatial averaging in seasonal simulations of the terrestrial water balance. *Hydrol Process Hyp*. <https://doi.org/10.1002/hyp.13745>
- Schertzer D, Lovejoy S (1987) Physical modeling and analysis of rain and clouds by anisotropic scaling multiplicative processes. *J Geophys Res* 92:9693. <https://doi.org/10.1029/JD092iD08p09693>
- Sharma A, Mehrotra R (2010) Rainfall generation. American Geophysical Union (AGU), Washington, pp 215–246
- Sivakumar B, Sharma A (2008) A cascade approach to continuous rainfall data generation at point locations. *Stoch Environ Res Risk Assess* 22:451–459. <https://doi.org/10.1007/s00477-007-0145-y>
- Subramanya K (1994) Engineering hydrology. Tata McGraw-Hill, Cambridge
- U.S. Department of Agriculture Soil Conservation Service, Urban Hydrology for Small Watersheds, Tech Release no. 55, June 1986
- Verhoest NEC, Vandenberghe S, Cabus P et al (2010) Are stochastic point rainfall models able to preserve extreme flood statistics? *Process* 24:3439–3445. <https://doi.org/10.1002/hyp.7867>
- Woolhiser DA, Osborn HB (1985) A stochastic model of dimensionless thunderstorm rainfall. *Water Resour Res* 21:511–522. <https://doi.org/10.1029/WR021i004p00511>
- Zeimet F, Schaeffli B, Artigue G et al (2018) Swiss rainfall mass curves and their influence on extreme flood simulation. *Water Resour Manag* 32:2625–2638. <https://doi.org/10.1007/s11269-018-1948-y>
- Zhang L, Singh VP (2007) Bivariate rainfall frequency distributions using Archimedean copulas. *J Hydrol* 332:93–109. <https://doi.org/10.1016/j.jhydrol.2006.06.033>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.