Stoch Environ Res Risk Assess (2006) 21: 195–212
DOI 10.1007/s00477-006-0056-3

ORIGINAL PAPER

Shiang-Jen Wu · Yeou-Koung Tung
Jinn-Chuang Yang

Stochastic generation of hourly rainstorm events

Published online: 9 June 2006

Abstract Occurrence of rainstorm events can be characterized by the number of events, storm duration, rainfall depth, inter-event time and temporal variation of rainfall within a rainstorm event. This paper presents a Monte-Carlo based stochastic hourly rainfall generation model considering correlated non-normal random rainstorm characteristics, as well as dependence of various rainstorm patterns on rainfall depth, duration, and season. The proposed model was verified by comparing the derived rainfall depth–duration–frequency relations from the simulated rainfall sequences with those from observed annual maximum rainfalls based on the hourly rainfall data at the Hong Kong Observatory over the period of 1884–1990. Through numerical experiments, the proposed model was found to be capable of capturing the essential statistical features of rainstorm characteristics and those of annual extreme rainstorm events according to the available data.

Keywords Rainstorm characteristics · Stochastic modeling · Monte Carlo simulation

1 Introduction

Rainfall data are often required in water-related engineering studies, such as flood forecast, prevention and mitigation, seepage and infiltration analysis for slope stability assessment. As a result, the quality and reliability of hydrosystem engineering design and analysis are affected by the length of rainfall record consisting of a sequence of rainstorm events. The rainstorm events can be characterized by the number of occurrence of rainstorm events and the associated depth, duration, inter-event (elapse) time as well as temporal pattern of rainfall hyetograph (rainstorm pattern) (Marien and Vandewiele 1986). In real-life hydrosystem design and analysis engineers are often faced with the problem of having insufficiently long rainfall record, especially when rainfall depth–duration–frequency (DDF) relationships are established on the basis of annual maximum data. Therefore, it would be desirable to develop practical and effective methods to fully utilize available rainfall records for maximum information extraction.

Several models have been developed to generate precipitation sequences and they can be broadly categorized into two types: (1) meteorologic models; (2) stochastic models (Onof et al. 2000). Meteorologic models are generally deterministic which produce precipitation and other weather events by a large and complex set of differential equations (Mason 1986). The stochastic models mainly take into account of spatial and temporal randomness of rainfall for modeling rainfall process. That is, stochastic models simulate rainstorm event sequences by using spatial and temporal statistical properties of rainfall process extracted from available records (Eagleson 1977; Waymire and Gupta 1981a, b, c; Lovejoy and Schertzer 1990; Gupta and Waymire 1994; Tan and Sia 1997; Guenni and Bardossy 2002).

According to the variables to be simulated, the stochastic rainfall modeling procedure can further be classified into two types according to: (1) rain-cell models and (2) rainstorm characteristics models. The concept behind the rain-cell models is that the rainfall sequences are composed of multiple rain cells. Waymire and Gupta (1981a, b, c) published a series of papers on the continuous time rainfall modeling which stimulated much of subsequent works. More recently efforts have been focused on the variation of two models, namely, the Bartlett–Lewis model and Nyman–Scott model. These two models have been applied to simulate point-rainfall

S.-J. Wu · J.-C. Yang
Department of Civil Engineering,
National Chiao Tung University, Hsinchu, Taiwan

Y.-K. Tung (✉)
Department of Civil Engineering,
Hong Kong University of Science and Technology,
Kowloon, Hong Kong
E-mail: cetung@ust.hk
Fax: +852-2358-1534

process (Waymire and Gupta 1981a, b, c; Cowpertwait 1991, 1994, 1998, 2004; Cowpertwait et al. 1996a, b; Glasbey et al. 1995; Verhoest et al. 1997; Onof et al. 2000; Koutsoyiannis and Xanthopoulos 1990; Koutsoyiannis 1994, 2001a, 2003; Koutsoyiannis and Manetas 1996; Koutsoyiannis and Onof 2001b)

Alternatively, rainstorm characteristics models generate rainstorm event sequences by preserving the statistical features of rainstorm characteristics, such as the number of occurrence of rainstorm events, the associated depth, duration, inter-event (elapse) time, and rainstorm pattern. Raudkivi and Lawgun (1970) proposed a first-order Markov process model along with a Weibull distribution to simulate rainstorm characteristics of duration, depth and time between events based on observed 10-min rainfall data. Eagleson (1977) considered the above three rainstorm characteristics of hourly rainfalls to be statistically independent and modeled them by simple exponential distributions. Marien and Vandewiele (1986) developed an at-site probabilistic rainfall generator considering independence of rainstorm duration, depth, and inter-arrival time using 10-min point rainfall data in Belgium. Acreman (1990) developed a stochastic model to generate hourly-based rainfall event sequences at the single-site in which the rainfall duration, total rainfall depth, dry spells are modeled by an exponential distribution, conditional gamma distribution and generalized Pareto distribution, respectively. Acreman's model further uses a beta distribution to define the rainfall patterns based on the average profile of observed rainstorm events. Lambert and Kuczera (1996) proposed a simple and parsimonious point-rainfall model capable of representing the inter-event time, storm duration, average event intensity and temporal distribution. They improved Eagleson's model by taking account of relationships between statistical moments of rainstorm characteristics. Haberlandt (1998) presented a renewal model to simulate dry-spell and wet-spell durations of rainstorm events and wet-spell amount by using Weibull and lognormal distributions, along with a relationship between intensity and wet-spell duration based on long-term (≥10 years) 5-min rain series.

As mentioned above, both rain-cell and rainstorm characteristics models generate rainfall events sequences by Monte Carlo simulation based on the probability distributions that fit the statistical features of rainstorm characteristics. In the Monte Carlo simulation, rainstorm characteristics have been assumed to be statistically independent (e.g., Eaglson's model, Bartlett–Lewis and Nyman–Scott model) or correlated (e.g., Lambert and Kuczera's model and Acreman's model). Their major difference is the means by which total rainfall depth of a specified duration is disaggregated into rainfalls of finer time resolution for establishing a rainfall hyetograph. Acreman's model uses a beta distribution to describe the averaged cumulated rainfall profile and Eagleson's model assumes the rainstorm patterns are rectangular pluses. In addition, the Bartlett–Lewis and Nyman–Scott models describe rainfall process as being consisted of separating and overlapping rain cells. Rainfall hyetographs then can be defined by the occurrences of a series of rain cells with varying, but constant, intensities of different durations which are allowed to overlap. To compare above models, Cameron et al. (2000) evaluated three stochastic rainfall models based on their ability to reproduce the standard and extreme statistics of 1- and 24-h seasonal maximum rainfall using observed hourly rainfall data at three sites in UK. The three models evaluated were the modified Eagleson's exponential model (MEEM), the cumulative density function and generalized Pareto distribution model (CDFGPDM) by Cameron et al. (1999), and the random parameter Bartlett–Lewis gamma model (RPBLGM). It was found that the MEEM and RPBLGM can effectively reproduce certain standard rainstorm statistics, but relatively poor in reproducing the statistics of 1 and 24-h seasonal maximum rainfall. Overall, the CDFGPDM generally performed well under all criteria.

The main thrust of this paper is to present a practical framework which stochastically generates correlated non-normal rainstorm characteristics, including rainfall depth ordinates defining the rainstorm pattern. The proposed model preserves the statistics of correlated rainstorm characteristics extracted from observed rainfall data. Unlike earlier works of assuming either statistical independence (e.g., Eaglson's model) or the simplistic relationships of statistical moments between the rainfall intensity and duration (e.g., Lambert and Kuczera's model), the proposed model employs a practical and flexible way to deal with multivariate non-normal rainstorm characteristics by incorporating their marginal distributions and correlations. Furthermore, the proposed model considers storms of varying patterns and their intrinsic variability to better capture what might be occurring in reality than adopting an averaged profile of all available rainstorm events.

2 Development of stochastic rainstorm event model

Occurrence of rainstorm events can be characterized by the number of events, storm duration, rainfall depth,

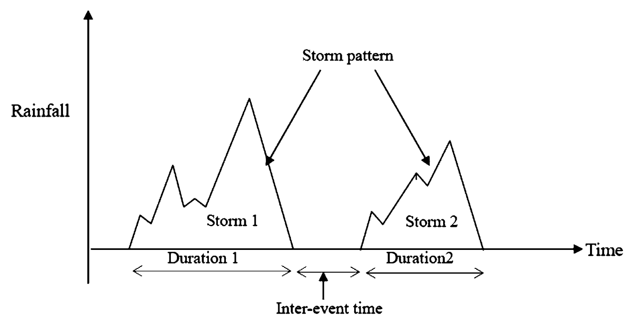

Fig. 1 Definition of rainstorm characteristics

inter-event time and rainstorm pattern as shown in Fig. 1. The stochastic model presented herein for generating rainstorm events is composed of three major components: (1) generation of the number of rainstorm events; (2) generation of storm depth, duration, and inter-event time for each event; and (3) generation of rainstorm pattern for each event.

2.1 Generation of number of rainstorm events

2.1.1 Definition of rainstorm event

To analyze the probabilistic properties of rainstorm characteristics and to synthesize rainstorm events, it is necessary to separate the time series of point-rainfall observations into individual events. From the physical and meteorological perspectives, it is desirable to treat individual rainstorm as an event within which rain may fall intermittently. In doing so, rainfall characteristics between different rainstorm events are assumed to be statistically independent for which several methods have been used to identify such rainstorm events: (a) autocorrelation method (Morris 1978); (b) rank correlation method (Bonta and Rao 1988); and (c) exponential method (Eagleson 1977; Bonta and Rao 1988; Restrepo and Eagleson 1982). The above methods do not base on the meteorologically meaningful feature; numerical investigations of rainfall records by the methods have also shown considerable degree of variability in determining rainstorm events.

Alternatively, many rainfall–runoff studies, particularly for stormwater drainage engineering, mainly concern with dry and wet periods of rainfalls composed of rainstorm events, regardless whether the rainfalls are from the same rainstorm or not. By this way of separating rainstorm events, two types of event are normally used: (1) events consisting of non-zero depths in rainfall sequences separated by dry time (Yen and Chow 1980) and (2) events consisting of a sequence that possibly contain zero depth according to the specific criteria.

It is expected that the simulated rainfall sequences based on the observed rainstorm characteristics would be affected by the adopted definition of rainstorm events. Since this study focuses on those rainstorm events that could potentially produce significant runoff, the criteria for defining and separating rainstorm events are based on the total rainfall amount, rainfall intensity, along with a specific threshold of dry period.

2.1.2 Modeling number of rainstorm events

To generate rainstorm sequences over a period of several years, the distribution properties for annual number of rainstorm events must be specified in advance. Poisson distribution is commonly adopted to describe annual random occurrence of hydrological events for the pulse-based and profile-based stochastic rainfall modeling (Eagleson 1977; Alexandersson 1985; Marien and Vandewiele 1986; Waymire and Gupta 1981a, b, c). Hence, Poisson distribution is considered and tested herein to model annual number of rainstorm event occurrences.

To test Poisson distribution, Cunnane (1979) applied the Fisher dispersion index, D_I, defined as

$$D_\mathrm{I} = \sum_{i=1}^{N} \frac{(m_i - \bar{m})^2}{\bar{m}} = \frac{(N-1)\mathrm{Var}(m)}{E(m)}, \tag{1}$$

where m_i is the annual number of hydrologic events observed in the ith year (i = 1, 2,..., N), and $\bar{m}$ is the sample mean of m_i. The test statistics D_I follows a χ^2 distribution with $N-1$ degree of freedom. The hypothesis on Poisson distribution is not rejected if the P-value associated with the sample D_I is larger than the specified significance level, which is set at 5 or 1% in general practice.

2.2 Generation of storm duration, rainfall depth, and inter-event time

In reality, storm duration, rainfall depth, and inter-event time are inherent correlated and their probabilistic distributions are likely to be non-normal. Therefore, the generation of these three rainstorm characteristics should preserve their respective marginal statistical properties and correlation relations.

2.2.1 Monte Carlo simulation for multivariate non-normal random variables

As mentioned above, storm duration, rainfall depth, and inter-event time could be a mixture of non-normal correlated variables and it is generally difficult to establish their joint distribution. To simulate multivariate non-normal random variates, Chang et al. (1997) proposed a practical procedure utilizing information on marginal distributions and correlations. The multivariate MCS procedure involves following three steps:

Transformation to standard normal space This step transforms correlated variables from their original domain to the standard normal space by the Nataf bivariate distribution model (Nataf 1962),

$$\rho_{ij} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{x_i - \mu_i}{\sigma_i}\right)\left(\frac{x_j - \mu_j}{\sigma_j}\right) \phi_{ij}(z_i, z_j | \rho_{ij}^*)\, \mathrm{d}z_i \mathrm{d}z_j, \tag{2}$$

where z_i and z_j are bivariate standard normal variables with the correlation coefficient ρ_{ij}^* and the joint standard normal density function $\phi_{ij}(\bullet)$; x_i and x_j are the correlated variables in the original space having, respectively, the means μ_i and μ_j, standard deviations σ_i and σ_j, and correlation coefficient ρ_{ij}. The equivalent correlation in the normal space ρ_{ij}^* can be obtained by solving Eq. 2 conditioned on the marginal PDFs of x_i and x_j as well as ρ_{ij}. Through an extensive numerical experiment, a set of semi-empirical formula has been developed by Liu and Der Kiureghian (1986) from which a transformation factor T_{ij}, depending on the marginal distributions and correlation of x_i and x_j, can be determined to modify ρ_{ij} in the original space to ρ_{ij}^* in the normal space by

$$\rho_{ij}^* = T_{ij} \times \rho_{ij} \tag{3}$$

Orthogonal transform After the normal transformation, orthogonal transform is applied to transform correlated multivariate standard normal variables into uncorrelated standard normal variables for which the random variates can be easily generated.

Inverse transformation Upon the generation of multivariate standard normal random variates z_i^*, one can obtain the corresponding random variates in the original space by the following inverse transformation

$$x_i = F_i^{-1}\left[\Phi\left(z_i^*\right)\right], \tag{4}$$

where $F_i(\cdot)$ is the marginal cumulative distribution function (CDF) of random variable x_i in the original space and $\Phi(\cdot)$ is standard normal CDF. The graphical procedure of the multivariate Monte Carol simulation is shown in Fig. 2.

2.2.2 *Marginal distributions for rainstorm duration, depth, and inter-event time*

According to Eq. 4, to invert the generated standard normal random variates back to the originally non-normal ones, the marginal distributions of rainstorm duration, depth, and inter-event time are to be known in advance. The best-fit marginal distribution of a random variable can be chosen by the Kolmogorov–Smirnov (K-S) test based on their sample statistics. In the K-S test, the probability distribution under consideration is not rejected if the sample value of the test statistic is less than the critical value associated with the desired significant level. This is equivalent to that the P-value corresponding to the sample test statistic value is larger than the desired significance level. However, it is common to encounter the situation that several probability distributions can be statistically plausible for a specified significance level. To choose one among several plausible distributions, this study selects the best-fit marginal distributions for the rainstorm duration, depth, and inter-event time that correspond to the largest P-value.

2.3 Generation of rainstorm pattern

A rainstorm hyetograph represents the time distribution of rainfall depth (or intensity) within a rainstorm event.

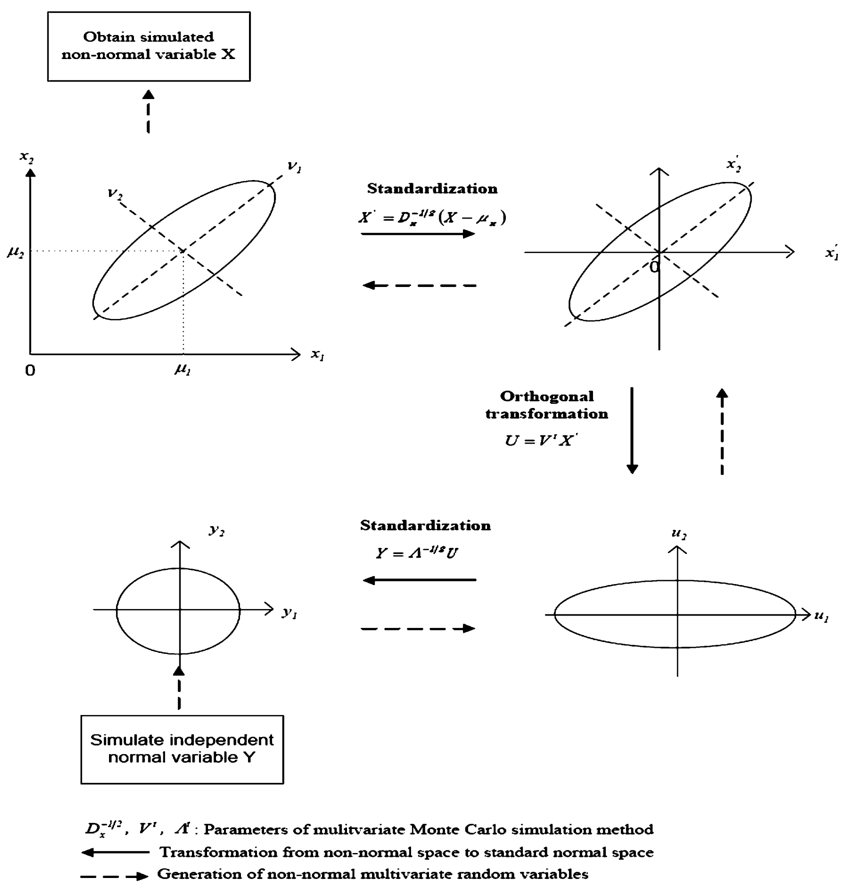

Fig. 2 Graphical illustration of multivariate Monte Carlo simulation (Tung and Yen 2005)

The rainstorm pattern (shape of hyetograph) has significant effects on the hydrologic response of a watershed. As rainstorm patterns vary from one event to another, it is necessary to have a procedure to generate the rainstorm patterns of different events in stochastic rainstorm simulation. In the following subsections, brief descriptions about the procedure are given and readers are referred to Wu et al. (2006) for more detail discussions.

2.3.1 Characterization of rainstorm patterns

During a rainstorm event, the rainfall intensities (depths) vary with time. Owing to the variation of rainstorm duration and total depth from one event to another, the classification of the similarity or dissimilarity of different rainstorm patterns can be best done by using a dimensionless scale. Non-dimensionlized rainstorm patterns can be obtained by adjusting the scale of the duration and depth of a rainfall mass curve as

$$\tau = \frac{t}{d}; F_\tau = \frac{D_{\tau\times d}}{D_d}; P_\tau = F_\tau - F_{\tau-1} \tag{5}$$

in which τ is the dimensionless time, $\tau \in (0, 1]$; d is the storm duration; F_τ is the dimensionless cumulative rainfall representing the cumulative fraction of rainfall depth, $F_\tau \in [0, 1]$; P_τ is the dimensionless incremental rainfall amount representing rainfall percentage increment in each time interval; D_t is the cumulative rainfall depth at time $t(t = \tau \times d)$; and D_d is the total rainfall depth. It is noted that F_τ at the first dimensionless time τ_1 is equal to P_{τ_1}. A graphical illustration of the non-dimensionalization of rainstorm pattern is shown in Fig. 3. The use of dimensionless rainstorm patterns removes the effects of storm duration and depth leaving the temporal variation as the sole factor differentiating different rainstorm patterns.

2.3.2 Identification of rainstorm pattern

As rainstorm pattern varies among events, it is practical to classify them into several representative types so that individual rainfall patterns within each type are similar to one another, but not necessarily identical, whereas individual rainfall patterns between different types are dissimilar. Non-dimensionalization allows rainstorm events with different durations and depths to be examined together to facilitate the identification of representative rainstorm patterns. This can be accomplished by statistical cluster analysis based on dimensionless rainfall mass curves. In this study, a non-dimensionalized rainfall mass curve is divided into 12 intervals and the corresponding dimensionless rainfall mass ordinates F_τ at $\tau = j/12$ with $j = 1, 2, ..., 12$ are used to represent a rainstorm pattern. In statistical cluster analysis, the K-mean clustering method on the basis of Euclidean distance (MacQueen 1967) is applied herein to categorize the patterns of all rainstorm events under consideration into several representative types.

2.3.3 Simulation of rainstorm pattern

After actual rainstorm patterns are classified into several distinct types, the simulation of rainstorm patterns for each event can be carried out in two steps: (1) generate rainstorm type and (2) generate plausible rainfall hyetographs for the rainstorm type under consideration. In step (1), the multinomial distribution is used to define the occurrence probabilities associated with each of the representative types of rainstorm pattern. As the occurrence of a particular rainstorm patterns might be affected by storm duration, depth, and season, such effects can be investigated by examining contingency tables shown later in model application section.

To generate rainfall hyetograph in step (2), two special features of dimensionless rainstorm patterns must be observed: (1) $P_{1/12} + P_{2/12} + \cdots + P_{12/12} = 1$ as $0 \le F_{1/12} \le F_{2/12} \le \cdots \le F_{12/12} = 1$ and (2) $0 \le P_\tau \le 1$ for $\tau = 1/12, 2/12, ..., 12/12$. In addition to the above two special features, the ordinates of the dimensionless rainfall hyetograph P_τ's are generally correlated non-normal random variables.

In this study, a practical approach using the log-ratio transformation (Aitchison 1986) is applied to treat $P_{1/12}$, $P_{2/12}$, ..., $P_{12/12}$ as compositional data. Furthermore, the Johnson distribution system is used to describe the marginal distribution of log-ratio variables associated with $P_{1/12}, P_{2/12}, ..., P_{12/12}$. The log-ratio transformation and Johnson distribution system are briefly described below.

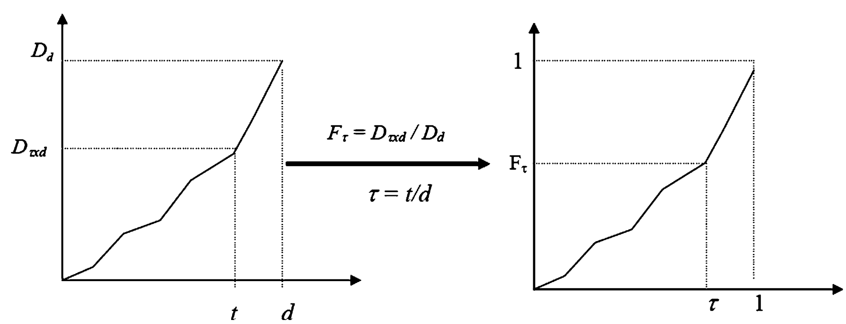

Fig. 3 Non-dimensionalization of storm profile

Log-ratio transformation To generate random constrained dimensionless rainfall hyetograph ordinates subject to $P_{1/12} + P_{2/12} + \cdots + P_{12/12} = 1$ and $0 \le P_\tau \le 1$, random $P_{1/12}, P_{2/12}, \ldots, P_{12/12}$ are transformed into a set of unconstrained variables by the following log-ratio transformation

$$R_\tau = \ln\left(\frac{P_\tau}{P_{\tau_*}}\right),\ \tau = 1/12,\ 2/12, \ldots 12/12;\ \tau \ne \tau_*, \tag{6}$$

where τ_* is the specified dimensionless time with P_{τ_*} being the associated non-negative dimensionless rainfall hyetograph ordinate.

Since $0 \le P_\tau \le 1$ for $\tau = 1/12, 2/12, \ldots, 12/12$, the transformed variables R_τ range from $-\infty$ to ∞. Note that, in the log-ratio transformation of actual data, neither P_τ nor P_{τ_*} can be zero to avoid numerical problem. The inverse transformation of log-ratio variables results in

$$P_\tau = \frac{\exp(R_\tau)}{1 + \sum\limits_{\substack{\tau' \ne \tau_* \\ \tau' = 1/12}}^{12/12} \exp(R_{\tau'})},\quad \tau = 1/12, 2/12, \ldots, 12/12 \tag{7}$$

$$P_\tau = \frac{1}{1 + \sum\limits_{\substack{\tau' \ne \tau_* \\ \tau' = 1/12}}^{12/12} \exp(R_{\tau'})},\quad \tau = \tau_*$$

As can be seen that $0 \le P_\tau \le 1$ and $P_{1/12} + P_{2/12} + \cdots + P_{12/12} = 1$ are automatically satisfied. In this study, generating rainstorm pattern is to simulate eleven correlated log-ratio variables R_τ of dimensionless rainfall ordinates, not involving the one at $\tau = \tau_*$.

Johnson distribution system for normal transformation To generate dimensionless rainfall hyetograph defined by the 11 log-ratio random variables R_τ's, it is necessary to identify the corresponding best-fit marginal distributions. This study adopted the Johnson distribution as the marginal distribution for R_τ. Fang and Tung (1996) found that the Johnson distribution system is more flexible to fit the rainstorm pattern than other distributions.

Johnson (1949) introduced a system of frequency curve consisting of four parameters

$$Z = \gamma + \delta\, g\left(\frac{X - \xi}{\lambda}\right) \tag{8}$$

where Z is standard normal random variable; $g(.)$ is a monotonic function of the original random variable X; ξ and λ are the location and scale parameters, respectively. There are three types of Johnson distribution: (1) lognormal system (S_L): $Z = \gamma + \delta \ln(X - \xi)$, $\xi \le X$ (2) unbounded system (S_U): $Z = \gamma + \delta \sinh^{-1}[(X - \xi)/\lambda]$; (3) bounded system ($S_B$): $Z = \gamma + \delta \ln[(X - \xi)/(\xi + \lambda - X)]$, $\xi \le X \le \xi + \lambda$.

These three curves cover the entire region of feasible distribution defined by the product–moment ratio diagram (Johnson 1949; Tadikamalla 1980). Hill et al. (1976) provided an algorithm to estimate the parameters of the Johnson distribution by matching the first four product–moments of X and to determine one of the best-fit Johnson distribution type.

2.3.4 *Procedure for generating dimensionless hyetograph*

Based on the characterization and identification of rainstorm events, log-ratio transformation, and MCS with Johnson distribution system, the dimensionless rainfall hyetograph ordinates P_τ are generated. Specifically, the generation procedure for P_τ is mostly similar to the procedure of generating the storm duration, depth and inter-event time shown in Fig. 2, but the only difference is that the Johnson distribution is selected to be candidate probability distributions. In addition, the parameters ($\gamma, \delta, \xi, \lambda$) in the Johnson distribution of each log-ratio variable R_τ associated with the rainstorm type under consideration are determined from the statistical moments of all log-ratio variables $R_{1/12}, R_{2/12}, \ldots, R_{12/12}$.

2.4 Generation of rainstorm events sequence

To simulate rainstorm sequences at a site, major steps in the proposed stochastic model are shown in Fig. 4 and are summarized below:

Step 1. Based on the retrieved rainstorm event data, summarize relevant seasonal statistical properties of rainstorm characteristics (number of events, duration, depth, inter-event time, ordinates of rainfall hyetograph, and type of rainstorm pattern) including their statistical moments, correlation relationships, and marginal probability distributions.

Step 2. Generate the number of rainstorm events for each season year-by-year over the simulated time period.

Step 3. For each generated rainstorm event, determine its rainstorm pattern type by the multinomial distribution and generate rainstorm depth, duration, and inter-event time by the procedure for simulating correlated multivariate non-normal variables described in Sect. 2.2.

Step 4. For each rainstorm event, according to its rainfall pattern type, generate the ordinates of dimensionless rainfall hyetograph using the log-ratio transformation along with the Johnson distribution described in Sect. 2.3.

Step 5. Convert generated dimensionless rainfall hyetograph to its original scale based on the corresponding depth and duration.

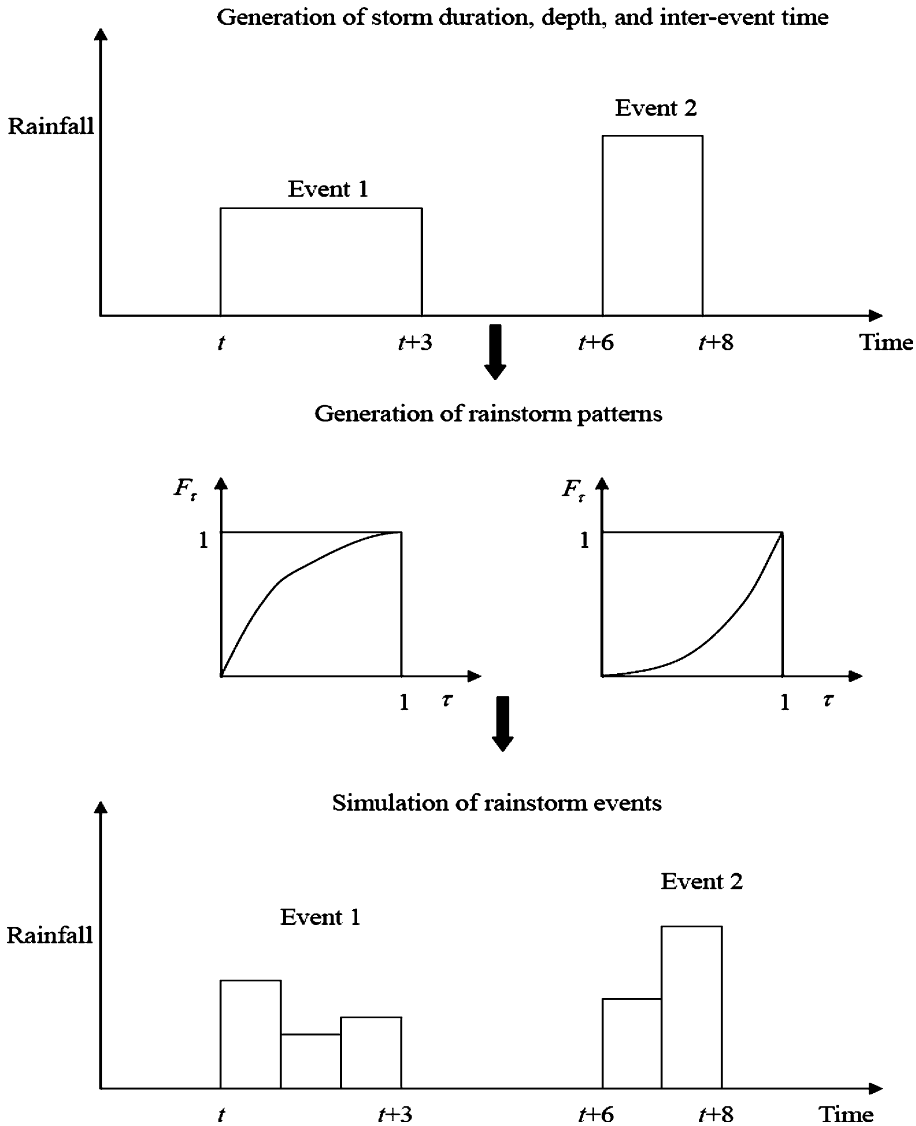

Fig. 4 Schematic diagram of generating rainstorm events

3 Model application

3.1 Description of rainfall data

For model application and performance evaluation, hourly rainfall data from 1884 to 1990 at the Hong Kong Observatory (HKO), with an interruption in 1940–1946 due to the World War II, were used to determine the distributional properties of rainstorm characteristics for the development of the stochastic rainstorm generation model

The criteria for extracting rainstorm events are based on dry-time, event rainfall amount, and hourly rainfall amount. As the study focuses rainfall events that could potentially produce significant runoff event, criteria adopted for rainstorm events retrieval are the followings: (1) dry-time ≥ 1-h; (2) event rainfall amount ≥ 30 mm/event; and (3) any hourly rainfall amount within an event ≥ 10 mm/h. According to the above criteria, a total of 1,690 rainstorm events over the period of 1884–1939, 1947–1990 were extracted and their associated rainstorm characteristics are analyzed.

3.2 Analysis of statistical features of storm characteristics

The number of occurrence of rainstorm events, storm duration, rainfall depth, inter-event time and temporal distribution of rainfall are the main features considered in the proposed stochastic model. To model rainstorm events more accurately, an understanding of these rainstorm characteristics as affected by seasonality is helpful through a statistical analysis of storm characteristics with respect to months or seasons.

3.2.1 Number of occurrence of rainstorm events

Table 1 shows the total number of rainstorm events over the record period by month, season, dry-period (Janu-

Table 1 Number of rainstorm event occurrences by month, season, dry- and wet-periods during 1884–1939, 1947–1990

Time period	Number of storm events
January	9
February	13
March	38
April	136
May	276
June	297
July	280
August	292
September	228
October	78
November	30
December	13
Spring (January–March)	60
Summer (April–June)	709
Autumn (July–September)	800
Winter (October–December)	121
Dry-period (Spring and Winter)	181
Wet-period (Summer and Autumn)	1,509
Total	1,690

ary–March and October–December), and wet-period (April–September). It is clearly observed that summer (April–June) and autumn (July–September) seasons have significantly higher number of rainstorm events than spring (January–March) and winter (October–December) seasons.

3.2.2 Storm duration, rainfall depth and inter-event time

By the two-sample K-S test for the equality of probability distributions, Table 2 shows that the *P*-value for storm duration, rainfall depth, and inter-event time in spring-winter and summer-autumn are greater than the 5% significant level. This indicates that the probability distributions for these storm characteristics in spring and winter seasons do not differ statistically. The same can be said about the storm characteristics in summer and autumn seasons. Therefore, according to Tables 1 and 2, it is plausible to divide a single year into dry- and wet-periods in which the former period contains winter and spring seasons while the latter includes summer and autumn seasons.

Table 2 *P*-values of two-sample K-S test on equality of distribution for storm duration, rainfall depth, and inter-event time in various seasons

Storm charac.	Season	Spring	Summer	Autumn
Storm duration	Summer	0.000		
	Autumn	0.000	0.092	
	Winter	0.152	0.000	0.000
Rainfall depth	Summer	0.000		
	Autumn	0.000	0.158	
	Winter	0.236	0.000	0.000
Inter-event time	Summer	0.000		
	Autumn	0.000	0.475	
	Winter	0.500	0.000	0.000

Table 3 Statistics of storm duration, rainfall depth, and inter-event time

Period	Mean	SD	Skew	Kurt	L-Cv	L-skew	L-kurt
Storm duration (mean and SD in hour)							
Dry	10.64	7.46	1.63	5.49	0.36	0.32	0.19
Wet	7.68	5.07	2.09	9.48	0.33	0.31	0.21
Year	8.00	5.46	2.12	9.19	0.34	0.32	0.21
Rainfall depth (mean and SD in mm)							
Dry	71.21	64.07	3.17	14.26	0.37	0.54	0.32
Wet	69.04	53.66	3.38	19.45	0.33	0.48	0.28
Year	69.28	54.89	3.37	18.74	0.34	0.48	0.29
Inter-event time (mean and SD in hour)							
Dry	398.08	604.48	1.66	5.05	0.73	0.50	0.15
Wet	271.8	341.33	2.19	9.82	0.61	0.40	0.16
Year	285.45	380.72	2.30	9.92	0.63	0.43	0.18

Table 3 summarizes statistical features, including product–moments and *L*-moments, derived from all available rainfall data in the dry- and wet-periods. The mean and standard deviation of storm duration, rainfall depth, and inter-event time in the dry-period are greater than those in the wet-period. This is because that significantly higher number of rainstorm events occurs in the wet-period (see Table 1) so that rainstorm events in the wet-period would have shorter duration with possibly higher rainfall intensity. Since the statistical properties of rainstorm characteristics in the dry-period and wet-period are very different, it is therefore sensible to develop a bi-seasonal stochastic model to capture seasonally varying characteristics of rainstorm events in the dry-periods and wet-periods. The need to consider seasonal variation in rainfall modeling was also pointed out by Haberlandt (1998).

As shown in Table 4, the storm duration and rainfall depth have strong positive correlation in both dry-periods and wet-periods. This implies that total rainfall amount for a rainstorm event has a general increasing tendency as the storm duration increases. However, Table 4 also reveals that the inter-event time is only weakly correlated with storm duration and rainfall amount. Such correlation features are incorporated in stochastic generation of rainstorm characteristics in dry- and wet-periods by the multivariate MCS described previously.

Table 4 Correlation coefficients of storm duration, rainfall depth, and inter-event time in different time periods

Period	Dur-Dep	Dur-IeT	Dep-IeT
Dry	0.685	−0.048	−0.159
Wet	0.662	0.025	−0.025
Year	0.656	0.025	−0.050

Dur Duration, *Dep* depth, *IeT* inter-event time

Table 5 Statistics and dispersion index for the number of storm event occurrences in dry- and wet-periods for first 20, 50, and 80 years of record

Record length	Period	Mean	SD	Dispersion index	*P*-value
First 20-year	Dry	1.60	1.43	24.22	0.188
	Wet	13.25	4.87	33.97	0.019
First 50-year	Dry	1.86	1.34	47.44	0.536
	Wet	13.70	4.61	76.01	0.008
First 80-year	Dry	1.68	1.26	75.23	0.599
	Wet	13.83	4.38	109.38	0.013
100-year	Dry	1.73	1.32	99.71	0.461
	Wet	14.25	4.45	137.58	0.006

To stochastically generate rainstorm events, the probability distribution for the number of rainstorm event occurrences should be determined in advance. Poisson process is often found suitable to describe the occurrences of random hydrologic events in time and space (Clark 1998). The sample values of dispersion index, Eq. 1, and the corresponding *P*-values for the three partial records and full 100-year record are listed in Table 5. It shows that in the dry-period the *P*-values are well above the normally specified significant levels of 1 and 5%. As in the wet-period, the *P*-values are lower than 5%, but higher than 1% except for the case of first $m = 50$-year. Overall speaking, the Poisson distribution is suitable to describe random occurrences of rainstorm events at the HKO in both dry- and wet-periods.

The results of goodness-of-fit test on the various probability distributions for storm depth, rainfall depth and inter-event time by the one-sample K-S test under different partial record lengths are summarized in Table 6. It is observed that the *P*-value associated with a given distribution for a particular rainstorm characteristic has a decreasing trend with increase in the record length. This is expected because as sample size increases it becomes easier to discern whether or not the hypothesized distribution is the true underlying distribution. There are more storm events in the wet-period than in the dry-period and the *P*-value for the former period is smaller than that of the latter period. Most *P*-values associated with candidate distributions considered for rainstorm characteristics are greater than 1% in the dry-period while those *P*-values for the wet-period of the first 50 and 80 years are close to zero. To select the best-fit distribution for each storm characteristic, the

Table 6 *P*-values of probability distributions for storm duration, rainfall depth, and inter-event time under different record lengths

Record length	Period	Storm duration		Rainfall depth		Inter-event time	
		Distribution	*P*-value	Distribution	*P*-value	Distribution	*P*-value
First 20-year	Dry	N	3.12E-02	N	9.50E-03	N	2.03E-01
		LN	3.62E-01	LN	1.19E-01	LN	2.37E-01
		GAM	2.54E-01	GAM	5.53E-02	GAM	2.05E-01
		EXP	2.00E-03	EXP	5.95E-02	EXP	7.22E-02
	Wet	N	0.00E+00	N	0.00E+00	N	0.00E+00
		LN	5.29E-02	LN	2.00E-04	LN	6.20E-03
		GAM	2.06E-02	GAM	0.00E+00	GAM	4.22E-01
		EXP	0.00E+00	EXP	0.00E+00	EXP	0.00E+00
First 50-year	Dry	N	5.00E-04	N	0.00E+00	N	1.67E-02
		LN	1.99E-01	LN	1.10E-02	LN	4.50E-02
		GAM	5.61E-02	GAM	3.00E-04	GAM	2.63E-01
		EXP	0.00E+00	EXP	3.00E-04	EXP	4.32E-02
	Wet	N	0.00E+00	N	0.00E+00	N	0.00E+00
		LN	6.00E-04	LN	0.00E+00	LN	0.00E+00
		GAM	1.00E-04	GAM	0.00E+00	GAM	3.20E-03
		EXP	0.00E+00	EXP	0.00E+00	EXP	0.00E+00
First 80-year	Dry	N	2.00E-04	N	0.00E+00	N	1.49E-02
		LN	1.95E-01	LN	7.00E-04	LN	8.50E-03
		GAM	6.06E-02	GAM	0.00E+00	GAM	1.06E-01
		EXP	0.00E+00	EXP	0.00E+00	EXP	9.33E-02
	Wet	N	0.00E+00	N	0.00E+00	N	0.00E+00
		LN	0.00E+00	LN	0.00E+00	LN	0.00E+00
		GAM	0.00E+00	GAM	0.00E+00	GAM	1.00E-04
		EXP	0.00E+00	EXP	0.00E+00	EXP	0.00E+00
100-year	Dry	N	0.00E+00	N	0.00E+00	N	2.20E-03
		LN	1.11E-01	LN	1.00E-04	LN	6.20E-03
		GAM	1.50E-02	GAM	0.00E+00	GAM	9.92E-02
		EXP	0.00E+00	EXP	0.00E+00	EXP	1.14E-02
	Wet	N	0.00E+00	N	0.00E+00	N	0.00E+00
		LN	0.00E+00	LN	0.00E+00	LN	0.00E+00
		GAM	0.00E+00	GAM	0.00E+00	GAM	0.00E+00
		EXP	0.00E+00	EXP	0.00E+00	EXP	0.00E+00

N Normal, *LN* lognormal, *GAM* gamma, *EXP* exponential one having the largest *P*-value is chosen. Hence, the best-fit distributions of storm characteristics in the dry-period and wet-period under various sub-record lengths can be obtained, except for those in wet-period of the first 50 and 80 years which are zero. Overall speaking, it is found that the best-fit distribution for storm duration and rainfall depth is the two-parameter lognormal distribution, while for inter-event time the two-parameter gamma distribution is the best-fit one. The same result for inter-event time was obtained by Yen et al. (1993) in their study on hourly rainfall at Urbana, IL using zero-rainfall period to separate rainstorm events. As a result of the K-S test, the lognormal distribution was employed in this study for generating storm duration and depth, whereas the gamma distribution was used for generating inter-event time.

3.2.3 Temporal distribution of rainfall (Rainstorm pattern)

Referring to Fig. 4, upon the generation of storm duration, rainfall depth and inter-event time, the next step is to generate plausible rainfall hyetographs to establish the sequence of rainstorm events. To achieve that, statistical features of dimensionless rainfall hyetograph ordinates defining rainstorm pattern at the HKO are examined. Statistical analysis for the temporal distributions of rainstorm patterns includes two parts: (1) identification of typical rainstorm patterns and (2) investigation of the factors affecting the frequency occurrence of rainstorm patterns.

Classification of typical rainstorm patterns In this study, typical rainstorm patterns are classified using the statistical cluster analysis based on the ordinates of dimensionless rainfall mass curve ordinates F_τ. According to the dimensionless mass curve of 1,690 rainstorm events extracted, six representative rainstorm patterns shown in Fig. 5 are identified through classification

Table 7 Contingency table of storm pattern based on storm duration and rainfall depth in wet-period (April–September)

Rainfall depth (mm)	Storm duration (h)				Total
	0–5	5–10	10–20	> 20	
Pattern-A1					
0–50	66	58	6	0	130
50–100	16	63	10	0	89
100–200	1	7	14	1	23
> 200	0	0	6	0	6
Total	83	128	36	1	248
Pattern-A2					
0–50	66	100	12	0	178
50–100	22	98	33	1	154
100–200	2	20	15	3	40
> 200	0	0	6	3	9
Total	90	218	66	7	381
Pattern-C					
0–50	55	92	10	0	157
50–100	19	56	21	0	96
100–200	2	17	23	5	47
> 200	0	0	5	2	7
Total	76	165	59	7	307
Pattern-U					
0–50	39	53	19	0	111
50–100	5	44	34	4	87
100–200	0	9	22	8	39
> 200	1	0	1	10	12
Total	45	106	76	22	249
Pattern-D1					
0–50	36	46	8	0	90
50–100	13	44	19	1	77
100–200	1	9	13	5	28
> 200	0	0	3	7	10
Total	50	99	43	13	205
Pattern-D2					
0–50	42	26	2	0	70
50–100	12	12	9	1	34
100–200	0	3	4	3	10
> 200	0	0	2	1	3
Total	54	41	17	5	117

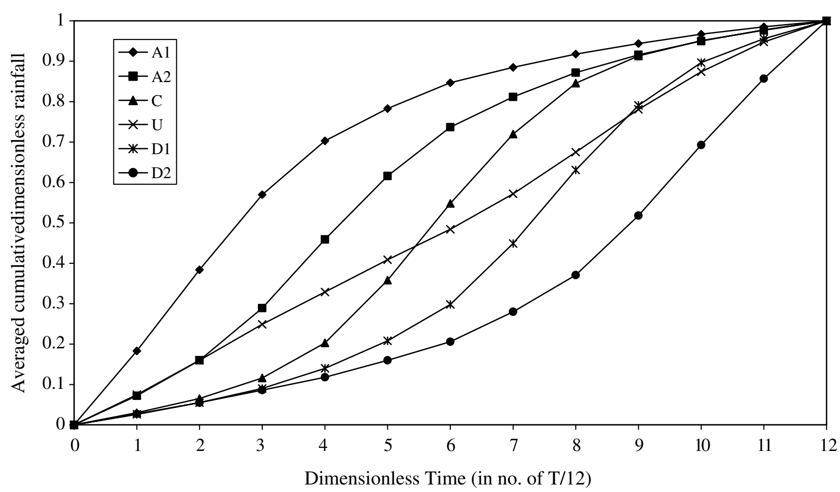

Fig. 5 Representative rainstorm patterns at the HKO

Table 8 Comparison of statistics of observed and simulated rainstorm characteristics

	Record	Duration (h)	Depth (mm)	Inter-event time (h)
Based on first 20-year observations[a]				
Mean	Entire	7.97	68.93	287.42
	Observed	8.15	72.26	293.09
	Simulated			
	Mean	8.17	72.36	301.75
	SD	0.13	1.41	8.26
SD	Entire	5.41	54.98	381.97
	Observed	5.44	60.22	406.84
	Simulated			
	Mean	5.45	60.25	405.59
	SD	0.26	2.99	17.04
Skew	Entire	2.08	3.44	2.29
	Observed	2.04	3.57	2.19
	Simulated			
	Mean	2.33	3.06	2.69
	SD	0.60	0.63	0.38
Kurtosis	Entire	8.87	19.34	9.94
	Observed	8.86	20.91	8.29
	Simulated			
	Mean	13.73	19.87	14.31
	SD	11.76	10.76	5.70
L-Cv	Entire	0.34	0.34	0.63
	Observed	0.33	0.35	0.65
	Simulated			
	Mean	0.33	0.39	0.61
	SD	0.011	0.01	0.01
L-skew	Entire	0.31	0.49	0.42
	Observed	0.31	0.49	0.46
	Simulated			
	Mean	0.30	0.36	0.45
	SD	0.03	0.02	0.02
L-kurt	Entire	0.21	0.29	0.18
	Observed	0.21	0.29	0.21
	Simulated			
	Mean	0.20	0.23	0.22
	SD	0.03	0.03	0.04

Correlation coefficient	Dur, Dep	Dur, IeT	Dep, IeT
Entire	0.66	0.04	−0.04
Observed	0.65	0.09	−0.04
Simulated			
Mean	0.66	0.09	−0.04
SD	0.02	0.03	0.02

	Record	Duration (h)	Depth (mm)	Inter-event time (h)
Based on first 50-year observations[b]				
Mean	Entire	7.97	68.85	287.42
	Observed	8.05	67.97	296.90
	Simulated			
	Mean	8.07	68.00	301.22
	SD	0.09	0.90	6.60
SD	Entire	5.41	54.98	381.97
	Observed	5.32	55.74	405.15
	Simulated			
	Mean	5.36	55.79	406.05
	SD	0.18	2.09	15.32
Skew	Entire	2.08	3.44	2.29
	Observed	2.13	3.92	2.21
	Simulated			
	Mean	2.23	3.39	2.50
	SD	0.44	0.54	0.30

Table 8 (Contd.)

	Record	Duration (h)	Depth (mm)	Inter-event time (h)
Kurt	Entire	8.87	19.34	9.95
	Observed	9.40	24.38	8.77
	Simulated			
	Mean	11.87	21.77	12.06
	SD	7.59	10.24	4.19
L-Cv	Entire	0.34	0.34	0.63
	Observed	0.33	0.33	0.65
	Simulated			
	Mean	0.33	0.36	0.63
	SD	0.01	0.01	0.01
L-skew	Entire	0.32	0.49	0.42
	Observed	0.31	0.50	0.44
	Simulated			
	Mean	0.31	0.49	0.46
	SD	0.02	0.01	0.01
L-kurt	Entire	0.21	0.29	0.18
	Observed	0.21	0.31	0.18
	Simulated			
	Mean	0.21	0.29	0.21
	SD	0.02	0.02	0.03

Correlation coefficient	Dur, Dep	Dur, IeT	Dep, IeT
Entire	0.66	0.04	−0.04
Observed	0.65	0.09	−0.04
Simulated			
Mean	0.66	0.09	−0.04
SD	0.02	0.03	0.02

Dur Duration, *Dep* depth, *IeT* inter-event time
[a]Entire = 100-year observed rainfall data, Observed = first 20-year observed rainfall data, Simulated = first 20-year observed + 80-year of simulated rainfall data
[b]Entire = 100-year observed rainfall data, Observed = first 50-year observed rainfall data, Simulated = first 50-year observed + 50-year of simulated rainfall data

process (Wu et al. 2006). The six rainstorm patterns consist of four basic types of hyetograph: advanced type (A1 and A2); central-peaked type (C); uniform type (U); and delayed type (D1 and D2). The advanced types have relatively higher rainfall intensity during early part of rainstorm event, whereas the delayed types are just the opposite. A central-peaked pattern has relatively high rainfall in the middle and tapers off towards the beginning and ending of the rainstorm event, while the uniform type reveals relatively constant rainfall intensity throughout the storm duration.

Factors affecting the occurrence of rainstorm patterns To investigate the frequency occurrence of rainstorm patterns affected by the storm duration, rainfall depth, and seasonality, the six representative rainstorm patterns are grouped by storm duration and rainfall depth in the form of contingency table (see Table 7 for wet-period as an example). It can be clearly observed that occurrence frequencies of the six representative rainstorm patterns are not uniform with respect to storm duration and rainfall amount at the HKO. Hence, occurrence frequencies of the six representative rain-

Table 9 Comparison of statistics of observed and simulated annual maximum rainfall for different durations

Statistics	Record length	1-h	2-h	6-h	12-h	24-h
Based on the first 20 years of observations[a]						
Mean (mm)	Entire	59.03	89.90	144.52	181.69	232.22
	Observed	53.57	85.19	150.92	187.85	250.69
	Simulated					
	Mean	52.98	86.47	151.07	186.96	217.82
	SD	1.51	2.39	4.01	4.92	5.71
SD (mm)	Entire	17.28	29.92	61.53	76.16	99.35
	Observed	15.69	33.91	72.25	83.26	130.01
	Simulated					
	Mean	19.22	32.02	59.36	76.00	95.32
	SD	1.94	2.81	5.43	7.20	8.37
Skew	Entire	0.70	0.68	1.97	1.58	1.48
	Observed	0.74	0.99	1.75	1.49	1.85
	Simulated					
	Mean	1.55	1.45	1.56	1.53	1.85
	SD	0.71	0.52	0.47	0.51	0.42
Kurt	Entire	3.20	3.17	8.77	7.03	6.94
	Observed	2.93	3.43	6.48	5.89	7.27
	Simulated					
	Mean	7.65	6.65	7.03	6.93	9.03
	SD	6.80	3.56	3.31	3.65	3.05
L-Cv	Entire	0.16	0.19	0.22	0.22	0.23
	Observed	0.17	0.23	0.25	0.24	0.27
	Simulated					
	Mean	0.19	0.20	0.21	0.21	0.22
	SD	0.03	0.03	0.01	0.01	0.03
L-skew	Entire	0.13	0.13	0.25	0.22	0.20
	Observed	0.20	0.23	0.29	0.23	0.27
	Simulated					
	Mean	0.22	0.22	0.22	0.22	0.24
	SD	0.10	0.11	0.04	0.04	0.09
L-kurt	Entire	0.12	0.12	0.22	0.16	0.16
	Observed	0.18	0.16	0.28	0.27	0.29
	Simulated					
	Mean	0.18	0.17	0.18	0.18	0.18
	SD	0.11	0.11	0.04	0.03	0.10
Based on the first 50 years of observations[b]						
Mean (mm)	Entire	59.03	89.90	144.52	181.69	232.22
	Observed	54.54	84.62	145.01	182.25	232.13
	Simulated					
	Mean	52.35	82.93	141.94	177.42	211.57
	SD	1.30	2.05	3.35	4.47	5.50
SD (mm)	Entire	17.28	29.92	61.53	76.16	99.35
	Observed	15.77	31.12	70.73	86.08	111.13
	Simulated					
	Mean	17.13	29.19	57.74	73.49	93.45
	SD	1.52	2.31	3.89	5.01	5.97
Skew	Entire	0.70	0.68	1.97	1.58	1.48
	Observed	0.76	1.04	2.18	1.96	1.93
	Simulated					
	Mean	1.20	1.28	2.11	1.86	1.97
	SD	0.41	0.42	0.34	0.33	0.31
Kurt	Entire	3.20	3.17	8.77	7.03	6.94
	Observed	3.25	3.87	8.75	7.80	7.99
	Simulated					
	Mean	5.41	5.46	10.11	8.47	9.32
	SD	2.59	2.87	2.53	2.20	1.99
L-Cv	Entire	0.16	0.19	0.22	0.22	0.23
	Observed	0.16	0.20	0.24	0.24	0.25
	Simulated					
	Mean	0.18	0.19	0.20	0.21	0.22
	SD	0.01	0.01	0.01	0.01	0.01
L-skew	Entire	0.13	0.13	0.25	0.22	0.20
	Observed	0.17	0.20	0.32	0.29	0.29
	Simulated					
	Mean	0.19	0.21	0.25	0.24	0.26
	SD	0.04	0.04	0.03	0.03	0.03
L-kurt	Entire	0.12	0.12	0.22	0.16	0.16
	Observed	0.13	0.16	0.27	0.23	0.19
	Simulated					
	Mean	0.16	0.17	0.22	0.20	0.19
	SD	0.03	0.03	0.03	0.03	0.02

[a]Entire = 100-year observed rainfall data, Observed = first 20-year observed rainfall data, Simulated = first 20-year observed + 80-year of simulated rainfall data
[b]Entire = 100-year observed rainfall data, Observed = first 50-year observed rainfall data, Simulated = first 50-year observed + 50-year of simulated rainfall data

storm patterns would depend on the storm duration and rainfall depth and such dependency should be accounted for in the simulation model for a more realistic generation of rainstorm events.

For a rainstorm event generated with the specified rainfall duration and depth in a dry or wet period, the random selection of a particular rainstorm pattern for that event can be made by using a multinomial distribution model with the following probability distribution,

$$Q_k = \frac{n_k}{\sum_{k'=1}^{6} n_{k'}}, k = 1, 2, \ldots, 6, \qquad (9)$$

where Q_k is the probability of occurrence of rainstorm pattern k; and n_k is the number of occurrences of rainstorm pattern k whose dependence on rainfall depth, duration, and season can be obtain from the contingency tables.

3.3 Performance evaluation and verification of proposed model

The proposed rainstorm generation model is evaluated by examining the general and extreme statistics of simulated rainstorm characteristics. In particular, on the extreme rainfall statistics, comparisons are made on the annual maximum rainfall frequency relations derived from the simulated rainstorm event sequence with those directly obtained on the basis of assumed ‘available’ rainfall data. The performance evaluation specifically aims at: (1) examining the ability of the proposed model to preserve the essential statistical features of relevant rainstorm characteristics and those of the annual maximum rainfall series and (2) investigating the capability of the proposed model to add synthesized rainfall record with the expectation to enhance the accuracy and reliability of rainfall frequency analysis. The evaluation procedure is outlined below.

Step 1. Based on the complete n-year rainfall data series calculate the statistical properties of rainstorm characteristics and annual maximum rainfall frequency relations, θ_n.

Step 2. To emulate the situation of adding synthesized rainfall record and to preserve its time-series structure, select the first m years out of the total record of n years ($m \leq n$) and treat them as the available data. Then, based on the 'available' rainfall data series, rainstorm events according to the prescribed criteria and annual maximum rainfalls of different durations are extracted. The rainstorm characteristics statistics and frequency quantiles from the m-year of 'available' rainstorm record is denoted as $\theta_{m,\,0}$. From the statistical characteristics of extracted rainstorm events in the first m years, apply the proposed model to simulate rainfall event sequences for a period of $(n - m)$ years. Calculate the statistics of simulated rainstorm characteristics and

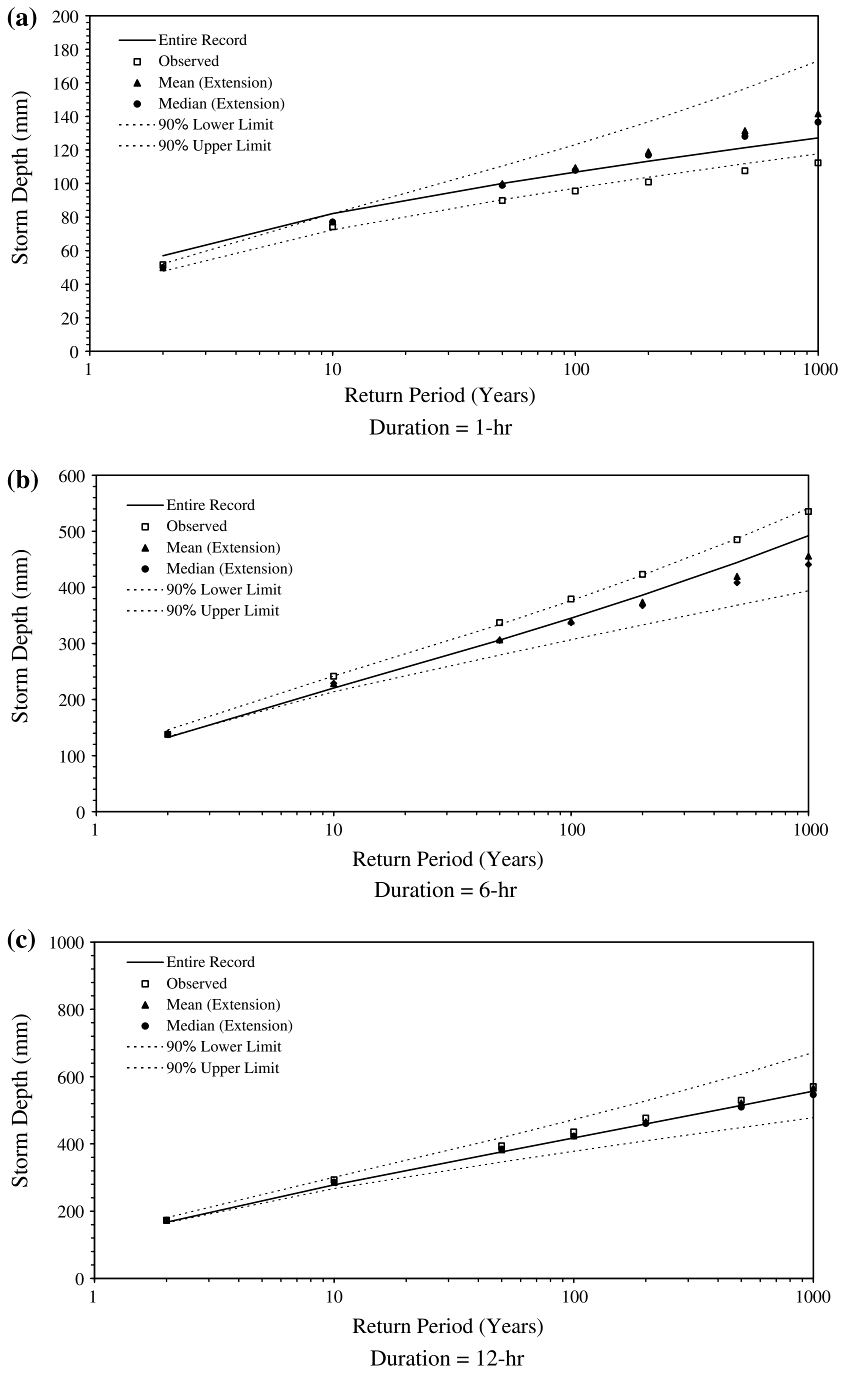

Fig. 6 Rainfall depth–frequency relations of various durations based on first $m = 20$ years of record. Duration = 1-h (**a**), 6-h (**b**), 12-h (**c**)

combine m-year of 'available' and $(n - m)$-year of simulated rainstorm events to estimate its frequency quantiles, denoted as $\theta_{m,\ n-m}$.

Step 3. Repeat step 2 a number of times to obtain the statistical properties of simulated rainstorm characteristics including their probability bounds.

Using 100 years of hourly rainfall record at the Hong Kong Observatory ($n = 100$ years), three partial record periods ($m = 20$, 50, and 80 years) starting from the beginning of the entire record (i.e., 1,884) were used as the 'available' record in the performance evaluation. Fifty synthesized rainstorm sequences of additional synthesized $(n - m)$-year are generated for each m-year of 'available' record.

Performance evaluation was conducted by comparing the statistical moments of rainstorm characteristics and those of annual maximum rainfalls of various durations from the simulated rainstorm events under the three

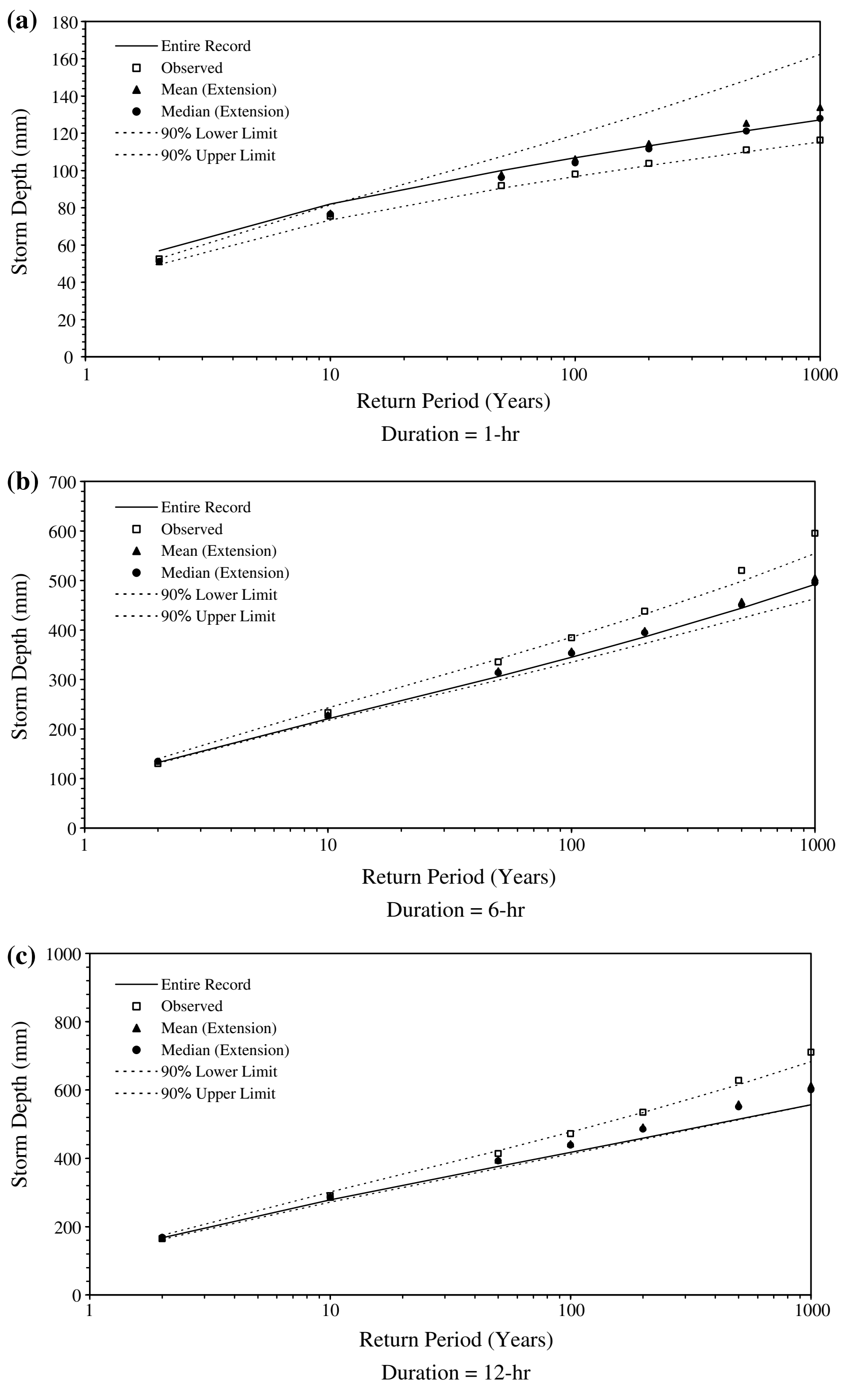

Fig. 7 Rainfall depth–frequency relations of various durations based on first $m = 50$ years of record. Duration = 1-h (**a**), 6-h (**b**), 12-h (**c**)

partial record lengths with those under the entire 100 years of observed rainfall data.

3.3.1 *Comparison of statistics of rainstorm characteristics*

For illustration, the first four product–moments and *L*-moments associated with the rainstorm characteristics of additional synthesized record sequences based on three sub-record lengths are compared with those from the full 100-year record. Table 8 shows the comparison of the first and second moments of observed and simulated rainstorm characteristics. It is observed that the mean values of the statistics of rainstorm characteristics with additional synthesized record are reasonably close with those from the full 100 years of observations (even with the first 20 years). The standard deviation associated with the simulation results reveals that the variation of rainstorm characteristics generated by the model and the values are relatively small, except for higher order product moments. Table 8 also indicate that as the partial record length increases, the discrepancy in statistical features of rainstorm characteristics between the full and partial records, as expected, decreases.

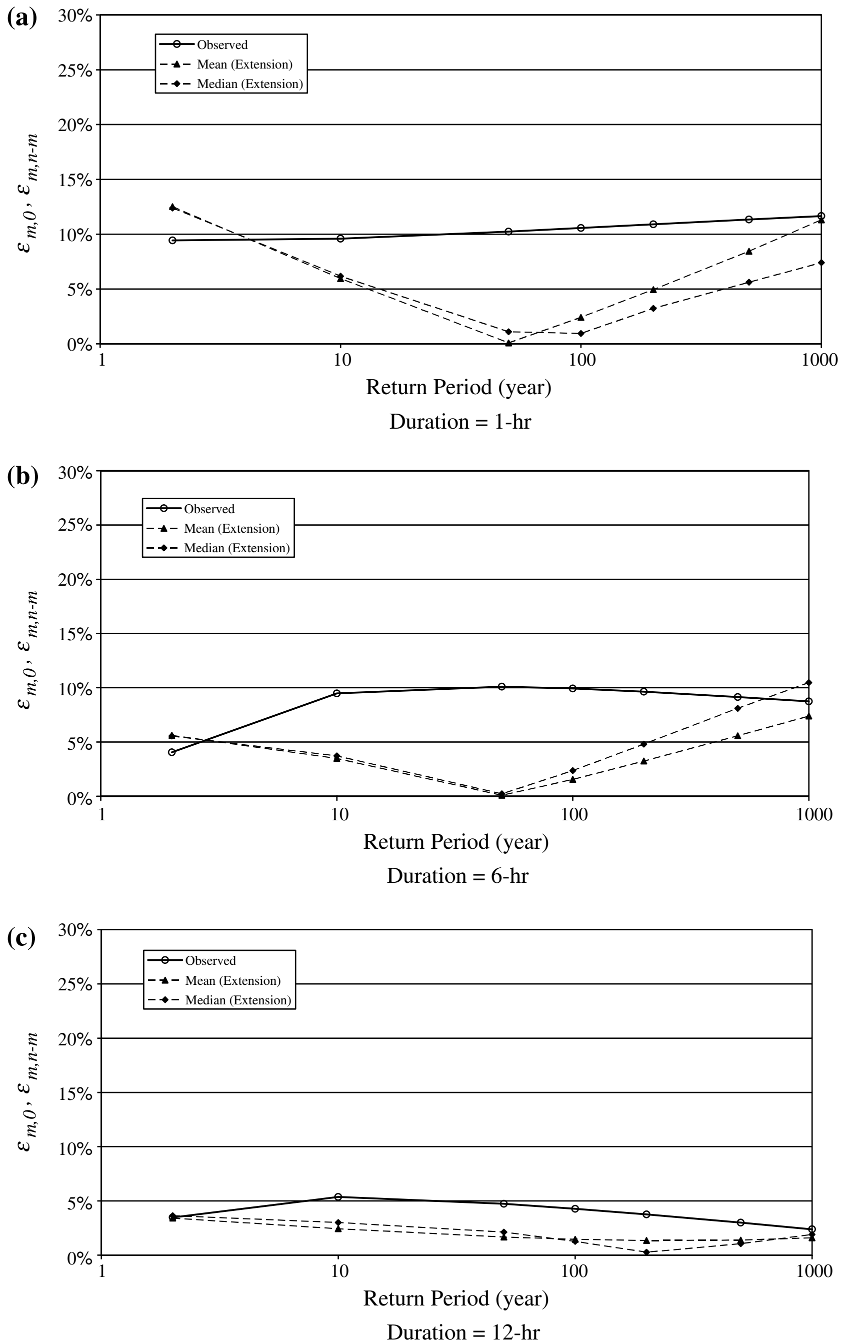

Fig. 8 Comparison of absolute error percentage in estimated rainfall depth–frequency curves of different durations (m = 20-years). Duration = 1-h (**a**), 6-h (**b**), 12-h (**c**)

3.3.2 Comparison of annual maximum rainfall statistics

Referring to Table 9, statistical moments associated with the annual maximum rainfalls of varying durations under different partial record lengths and data samples are listed. It can be observed that the third and fourth-order moments (especially the product–moments) of the annual maximum rainfalls from the data series with additional synthesized data are not close to those of the entire data as the first two moments. Between the data series with and without additional synthesized data the former offers a closer match for the second moments to those of the 100-year data series, especially when the partial record length is short and the storm duration is longer. For 1-h annual maximum rainfall, the proposed model provides less accurate estimation of the statistical moments than those without additional synthesized data. This could be attributed to the limitation of simulated hourly rainfall profile to capture the true temporal pattern of the short-duration rainstorm events.

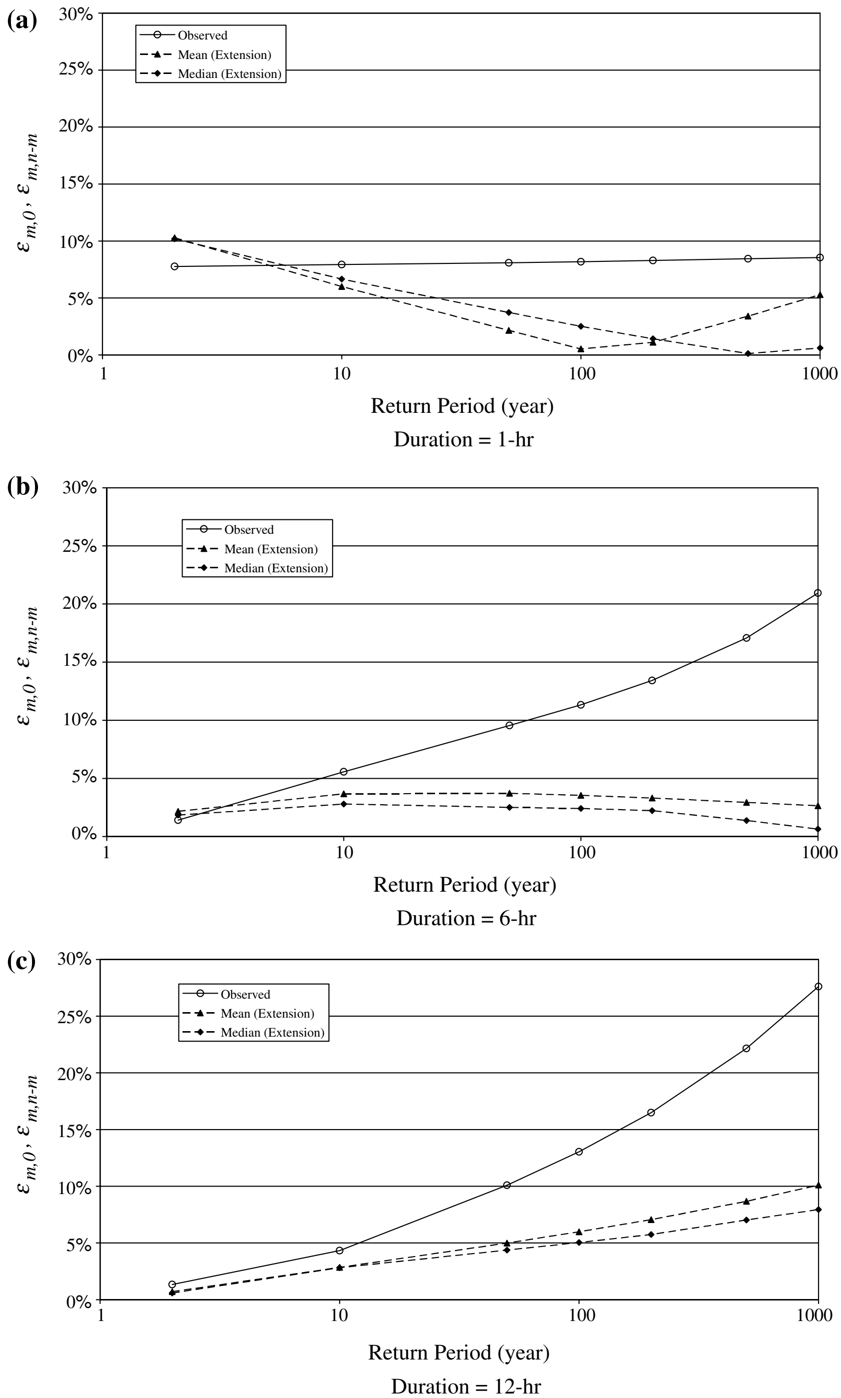

Fig. 9 Comparison of absolute error percentage in estimated rainfall depth–frequency curves of different durations (m = 50-years). Duration = 1-h (**a**), 6-h (**b**), 12-h (**c**)

From the well-known frequency factor equation, i.e., $X_T = \mu + \sigma K_T$, one can see that the frequency relationship is a combined effect of statistical moments of various orders, it may not be so easy to assess how the accuracy of individual moment would affect the overall behavior of the entire frequency curve.

3.3.3 Comparison of annual maximum rainfall frequency relationship

Rainfall depth–frequency relations derived from the 'available' partial record and those using additional synthesized data are compared with the depth–frequency relations from the complete set of 100-year. For illustration, Figs. 6 and 7, respectively, show the mean, medium, 90% bound of rainfall depth–frequency curves of three selected storm durations (1-, 6-, and 12-h) derived under 'available' record length of 20 and 50 years. Note that 90% bound indicates the random variability of outputs from the proposed model. It is observed that the 90% bound of simulated depth–frequency curve can capture the 'true' frequency curve based on the entire 100-year complete record. Furthermore, by adding synthesized data to the in 'available' record, the mean and median of simulated depth–frequency curves are closer to the 100-year curve and the 90% bound becomes tighter. It is also interesting to observe that, based solely on the 'available' data without adding synthesized data, the depth–frequency curves for some durations (i.e., 6 and 12-h based 20 and 50 years of 'available' data) lie outside the 90% bound of simulated results.

From Figs. 6 and 7 one observes that the mean depth–frequency curve lies relatively higher above the median curve indicating that the proposed stochastic rainstorm generating model produces positively-skewed estimation of rainfall quantile, especially for larger return period, such as 200-year or above, and shorter storm duration (1- or 2-h) when the 'available' record length is 20 or 50 years. As the storm duration and 'available' record length become longer, the mean and median depth–frequency curves coincide together. For the majority of the cases considered, the median depth–frequency curves are closer to those obtained on the basis of full 100-year record.

Figures 8 and 9 show the absolute error percentage of the depth–frequency relationships, defined as

$$\varepsilon_{m,0} = \left| \frac{\theta_{m,0} - \theta_n}{\theta_n} \right| \times 100\%; \varepsilon_{m,n-m} = \left| \frac{\theta_{m,n-m} - \theta_n}{\theta_n} \right| \times 100\% \quad (10)$$

under different partial record lengths with respect to those established by using full 100-year records in which $\theta = x_T{}^t$ denoting estimated T-year t-h rainfall depth. It is clearly observed that both median and mean depth–frequency curves calculated from combining 'available' and synthesized data using the proposed rainstorm generation model would enhance the accuracy of rainfall frequency analysis, especially when the return period is high.

Numerical results also point out that if the return periods of interest is significantly smaller than 'available' record length [e.g., $T \leq$ 5-year with m = 20-year; $T \leq$ 10-year with m = 50-year; $T \leq$ 50-year with m = 80-year (not shown)], adding synthesized record is a futile task which produces no improvement on the estimated quantiles, if not making the estimates less accurate. On the other hand, it is interesting to note that, for the limited conditions considered herein, the addition of rainstorm record by the proposed model has good potential to greatly enhance the accuracy of estimated rainfall quantiles with return period closer to or larger than the 'available' record length.

4 Conclusions

This paper presents a practical stochastic model for generating hourly-based rainstorm events according to the statistics of rainstorm characteristics, including number of occurrence of rainstorm events, storm duration, rainfall depth, inter-event time and rainstorm pattern. The proposed rainstorm generation model involves the use of Poisson distribution for generating number of rainstorm events and multivariate Monte Carlo simulation for storm duration, rainfall depth, and inter-event time, as well as constrained multivariate simulation conditioned on the total rainfall amount and storm duration for the corresponding rainfall hyetograph.

From the numerical experiments, the proposed model was found to be capable of capturing the essential statistical features of rainstorm characteristics and extreme events based on the available data. Furthermore, the proposed model shows promising potential to improve the accuracy of short rainfall record in establishing rainfall IDF relations, especially for events having return period close to or longer than the available record length.

Note that, in this study, the simulations of rainstorm events were made for the dry- and wet-periods in the example application. The proposed model, however, can be applied to deal with rainfall data associated with any time-interval of interest, such as month, if necessary. Moreover, as the proposed model requires the specification of statistical properties of rainstorm characteristics at a site, it is applicable for simulating rainstorm events at ungauged sites provided that relevant information about rainstorm characteristics are estimated by a proper regional analysis.

Acknowledgments This study is conducted under the auspice of research project "HKUST6016/01E: Investigating Issues in Rainfall Intensity-Duration and Time-Scale Relations in Hong Kong" funded by the Research Grant Council of Hong Kong Special Administration Region.

References

Acreman MC (1990) A simple stochastic model of hourly rainfall for Farnborough, England. Hydrol Sci J 35(2):119–148

Aitchison J (1986) Statistical analysis of compositional data. Chapman and Hall Inc., New York

Alexandersson H (1985) A simple stochastic model of the precipitation process. J Clim Appl Meteorol 24:1285–1295

Bonta JV, Rao AR (1988) Factors affecting the identification of independent rainstorm events. J Hydrol 98:275–293

Cameron DS, Beven KJ, Tawn J, Blazkova S, Naden P (1999) Flood frequency estimation for a gauged upland catchment (with uncertainty). J Hydrol 219:169–187

Cameron DS, Beven K, Tawn J (2000) An evaluation of three stochastic rainfall models. J Hydrol 228:130–149

Chang CH, Yang JC, Tung YK (1997) Incorporate marginal distributions in point estimate methods for uncertainty analysis. J Hydrol Eng 123(3):244–251

Clark RT (1998) Stochastic processes for water scientists: development and application. Wiley, New York

Cowpertwait PSP (1991) Further evelopment of the Neyman–Scott clustered point process for modelling rainfall. Water Resour Res 27(7):1431–1438

Cowpertwait PSP (1994) A generalized point process model for rainfall. Proc R Soc Lond A 447:23–37

Cowpertwait PSP (1998) A Poisson-cluster model of rainfall: high-order moments and extreme values. Proc R Soc Lond A 457:885–898

Cowpertwait PSP (2004) Mixed rectangular pulses models for rainfall. Hydrol Earth Syst Sci 8(5):993–1000

Cowpertwait PSP, O'Connell PE, Metcalfe AV, Mawdsley JA (1996a) Stochastic point process modelling of rainfall. I. Single-site fitting and validation. J Hydrol 175:17–46

Cowpertwait PSP, O'Connell PE, Metcalfe AV, Mawdsley JA (1996b) Stochastic point process modelling of rainfall. II. Regionalisation and disaggregation. J Hydrol 175:47–65

Cunnane C (1979) A note on the Poisson assumption in partial duration series modes. Water Resour Res 15(2):489–494

Eagleson PS (1977) The distribution of annual precipitation derived from observed storm sequence. Water Resour Res 14(5):713–721

Fang TQ, Tung YK (1996) Analysis of Wyoming extreme precipitation patterns and their uncertainty for safety evaluation of hydraulic structure. Technical Report, WWRC-96.5, Wyoming Water Resource Center, University of Wyoming, Laramie, Wyoming

Glasbey CA, Cooper G, McGechan MB (1995) Disaggregation of daily rainfall by conditional simulation from a point-process model. J Hydrol 165:1–9

Guenni L, Bardossy A (2002) A two step disaggregation method for highly seasonal monthly rainfall. Stoch Environ Res Risk Assess 16:188–206

Gupta VK, Waymire EC (1994) A statistical analysis of mesoscale rainfall as a random cascade. J Appl Meteor 32:251–267

Haberlandt U (1998) Stochastic rainfall synthesis using regionalized model arameters. J Hydrol Eng 3(3):160–168

Hill ID, Hill R, Holder RL (1976) Algorithm AS 99 fitting Johnson curves by moments. Appl Statist 25:180–189

Johnson NL (1949) System of frequency curves generated $z_i = \gamma + \delta \ln [y^{\alpha}{}_i / (1 - y^{\alpha}{}_i)]$ by method of translation. Biometrika 36:149–176

Koutsoyiannis D (1994) A stochastic disaggregation method for design storm and flood synthesis. J Hydrol 156:193–225

Koutsoyiannis D (2001a) Coupling stochastic models of different timescales. Water Resour Res 37(2):379–391

Koutsoyiannis D (2003) Multivariate rainfall disaggregation at a fine timescale. Water Resour Res 39(7):1173–1190

Koutsoyiannis D, Manetas A (1996) Simple disaggregation by accurate adjusting procedures. Water Resour Res 32(7):2105–2117

Koutsoyiannis D, Onof C (2001b) Rainfall disaggregation using adjusting procedures on a Poisson cluster model. J Hydrol 246:109–122

Koutsoyiannis D, Xanthopoulos T (1990) A dynamic model for short-scale rainfall disaggregation. Hydrol Sci J 35(3):303–322

Lambert M, Kuczera G (1996) A stochastic model of rainfall and temporal patterns. In: Proceedings on stochastic hydraulics. Balkema, Rotterdam

Liu PL, Der Kiureghian A (1986) Multivariate distribution models with prescribed marginal covariances. Probab Eng Mech 1(2):105–112

Lovejoy S, Schertzer D (1990) Multifractal, universality classes, and satellite and radar measurements of cloud rain fields. J Geophys Res 95:2021–2031

MacQueen J (1967) Some methods for classification and analysis of multivariate observations. In: Proceedings of the 5th Berkeley symposium, vol 1, pp 281–297

Marien JL, Vandewiele GL (1986) A point rainfall generator with internal storm structure. Water Resour Res 22(4):475–482

Mason J (1986) Numerical weather prediction. Proc R Soc Lond A407:51–60

Morris CD (1978) A stochastic model for an intermittent hydrologic process. PhD Thesis, University of Illinois at Urbana-Champaign

Nataf A (1962) Determinaiton des distributions don't les marges sont donnees. Computes Rendus de l'Academie des Sciences Paris 225:42–43

Onof C, Chandler RE, Kakou A, Northrop P, Wheater HS, Isham V (2000) Rainfall modelling using Poisson-cluster process: a review of developments. Stoch Environ Res Risk Assess 14:384–411

Raudkivi AJ, Lawgun N (1970) Synthesis of urban rainfall. Water Resour Res 6(2):445–464

Restrepo PJ, Eagleson PS (1982) Identification of independent rainstorms. J Hydrol 55:303–319

Tadikamalla PR (1980) On simulation non-normal distributions. Psychometrika 45(2):273–279

Tung YK, Yen BC (2005) Hydrosystem engineering uncertainty analysis. McGraw-Hill Book Company, New York

Tan SK, Sia SY (1997) Synthetic generation of tropical rainfall time series using an event-based method. J Hydro Eng 2(2):83–99

Verhoest N, Troch PA, De Troch FP (1997) On the applicability of Bartlett–Lewis rectangular pulses models in the modeling of design storms at a point. J Hydrol 202:108–120

Waymire E, Gupta VK (1981a) The mathematical structure of rainfall representation. 1. A review of stochastic rainfall models. Water Resour Res 17(5):1261–1272

Waymire E, Gupta VK (1981b) The mathematical structure of rainfall representation. 2. A review of the theory of point processes. Water Resour Res 17(5):1273–1285

Waymire E, Gupta VK (1981c) The mathematical structure of rainfall representation. 3. Some applications of the point process theory to rainfall processes. Water Resour Res 17(5):1286–1294

Wu SJ, Tung YK, Yang JC (2006) Identification and stochastic generation of representative rainfall temporal patterns in Hong Kong territory. Stoch Environ Res Risk Assess 20(3):171–183

Yen BC, Chow VT (1980) Design hyetographs for small drainage structures. J Hydraul Eng ASCE 106(HY6):1055–1076

Yen BC, Riggins R, Ellerbroek JW III (1993) Probabilistic characteristics of elapsed time between rainfalls. Management of irrigation and drainage systems: integrated perspectives, ASCE, pp 424–431