ORIGINAL PAPER

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

Identification and stochastic generation of representative rainfall temporal patterns in Hong Kong territory

Abstract In hydrosystem engineering design and analysis, temporal pattern for rainfall events of interest is often required. In this paper, statistical cluster analysis of dimensionless rainfall pattern is applied to identify representative temporal rainfall patterns typically occurred in Hong Kong Territory. For purpose of selecting an appropriate rainfall pattern in engineering applications, factors affecting the occurrence of different rainfall patterns are examined by statistical contingency tables analysis through which the inter-dependence of the occurrence frequency of rainfall patterns with respect to geographical location, rainfall duration and depth, and seasonality is investigated. Furthermore, due to inherent variability of rainfall mass curves or hyetographs within each classified rainfall pattern, a practical procedure to probabilistically generate plausible rainfall patterns is described. The procedure preserves the inherent stochastic features of random dimensionless rainfall hyetograph ordinates, which in general are correlated non-normal multivariate compositional variables.

Keywords Rainfall pattern · Cluster analysis · Johnson distribution system · Contingency table · Constrained Monte–Carlo simulation

Introduction

Information about rainfall events is often required in the design, analysis, and performance evaluation of a

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,
Clear Water Bay, Kowloon, Hong Kong
Fax: +852-2358-1534
E-mail: cetung@ust.hk

hydrosystem. Characteristics of a rainfall event are describable by its total depth, duration, and temporal variation of rainfall intensity within the event, which could influence the design of hydrosystem infrastructures. For example, design inflow hydrographs of selected frequencies required in dam safety evaluation are often computed from a synthetic (design) rainfall events in conjunction with the use of a proper rainfall–runoff model.

Various methods have been developed to establish synthetic rainfall events, which can be categorized into following five types: (1) by subjective judgment (Ogrosky[1964]); (2) by way of an adopted rainfall intensity–duration–frequency (IDF) relationship along with an assumed temporal pattern (US Army Corp of Engineers [1948]; Keifer and Chu[1957]; Bandyopadhyay[1972]; Chen[1976]); (3) by relative ranking of rainfall hyetograph ordinates (Pilgrim and Cordery[1975]); (4) by rainfall mass curve (Hershfield[1962]; Huff [1967]; Soil Conservation Service (SCS) [1972]); and (5) by the statistical moments of rainfall hyetograph (Yen and Chow [1980], [1983]). Rainfall IDF-based methods, which focus on intense bursts within rainfall events, are based on the maximum rainfall for a given duration from different rainfall events, rather than from a particular event. The resulting rainfall pattern is an envelope curve, which tends to be conservative for general design of a hydrosystem infrastructure. The shapes of IDF-based rainfall hyetographs, in general, do not resemble the rainfall patterns that actually occur. For the remaining types of methods, rainfall patterns are primarily established through empirical analysis of actual rainfall events. A detailed review of the methods for establishing design rainfall patterns can be found elsewhere (Yen and Chow [1983]). Often times, averaging process is applied to rainfall events of varying patterns without prior classification. The incorporation of rainfall pattern information has been shown to improve the current practice in stochastic generation of rainfall events (Lambert and Kuczera[1996]; Heneker et al. [2001]) in which the rainfall events are normally represented by rectangular pulses of

various durations and intensities (e.g., Onof and Wheater1993).

For the design and analysis of hydrosystems using synthetic rainfall hyetographs, improvement can be made from choosing appropriate rainfall pattern of possible types that have occurred. As each actual rainfall event is unique with respect to its duration, depth, and pattern it is practical to be able to identify a few representative rainfall patterns according to the shape of actual rainfall events. One of the objectives of this study is to identify and categorize event-based hourly rainfall patterns typically occurred in Hong Kong Territory by statistical cluster analysis. Furthermore, realizing the existence of inherent variability in rainfall pattern of individual event within each type, the study describes a practical procedure to stochastically generate plausible rainfall hyetographs of the specified pattern for hydrosystem engineering applications.

Characterization of rainfall patterns

During a rainfall event, rainfall intensity varies with time. As rainfall duration and total depth vary from one event to another, characterization of similarity or dissimilarity of different rainfall patterns can best be made through the use of a dimensionless scale.

Non-dimensionalization of rainfall pattern can be achieved by dividing the cumulative rainfall depths at different times D_t by the total rainfall depth D_T as $F_\tau = D_t / D_T$ and the time t by the rainfall duration (T) as $\tau = t/T$. By using dimensionless rainfall mass curves or hyetographs, the effects of rainfall duration and depth are removed, leaving the temporal variation of rainfall event as the only factor for differentiating events of different patterns. After non-dimensionalization, rainfall events with different depths and durations can be combined, examined, and categorized for identifying representative rainfall patterns.

A dimensionless rainfall mass curve shows the cumulative fractions of rainfall depth, $0 \leq F_\tau \leq 1$, over the dimensionless time, $0 \leq \tau \leq 1$. To characterize the temporal pattern of a rainfall event, several time points within the rainfall duration must be selected to extract the corresponding rainfall mass curve ordinates. Note that too few points might not accurately describe the underlying rainfall pattern whereas, on the other hand, too many points would capture unnecessary fine details that might mask the essential time-varying feature of the rainfall pattern. In this study, the entire duration of rainfall event is divided into 12 equal intervals ($\Delta T = T/12$) and the corresponding dimensionless rainfall mass curve ordinates F_τ or hyetograph ordinates $P_\tau = F_\tau - F_{\tau-1}$ at $\tau = k/12$, $k = 1, 2,..., 12$, are extracted to define the rainfall pattern.

To characterize rainfall pattern there are methods which parametrically treat the dimensionless rainfall mass curve as a cumulative distribution function (CDF) of the dimensionless time, τ. Hence, the temporal pattern of rainfall events can be characterized by the statistical moments of dimensionless time (Yen and Chow 1980; Fang and Tung 1996). The dimensionless cumulative rainfall mass curves can parametrically be fitted by a suitable CDF model, such as the β distribution (Acreman1990) and the Johnson distribution system (Stuart and Ord1987; Fang and Tung 1996).

The advantage of describing a rainfall pattern by its statistical moments or an analytical probability distribution function is to use small number of attributes to define the shape of a rainfall pattern. However, this advantage could also be a handicap for its inability to differentiate subtle difference among different rainfall patterns that might be important for a particular application due to the smoothing effect of adopting a parametric distribution function. According to a study on Wyoming hourly rainfall data (Fang and Tung 1996), it was found that the use of a 4-parameter Johnson distribution system, as expected, fits the rainfall pattern better than a 2-parameter β distribution.

Database for Hong Kong rainfall events

Hourly rainfall data from 16 automatic raingauges in Hong Kong Territory shown in Fig. 1 are used to extract pertinent rainfall events for analysis. The station number, location, elevation, and record length of each station are listed in Table 1. In data retrieval, the key decision is the definition based on which a rainfall event is extracted. Yen and Chow (1983) as well as Yen et al. (1993) reviewed various ways of defining a rainfall event for hydrological engineering applications from different perspectives. As watersheds in Hong Kong Territory are generally small and many are situated in mountainous terrains, the flow time-of-concentration is relatively short (in the order of 1–2 h). Hence, this study defines that two consecutive rainfall events are separated by a dry spell of more than 2 h. Also, as the study focuses on potentially flood-causing rainfalls, events with a total rainfall depth larger than 50 mm or hourly rainfall intensity larger than 10 mm/h in any hour within the event will be extracted.

Since the hourly based rainfall duration is not necessarily equal to the multiples of 12, the values of rainfall mass curve ordinates, F_τ at $\tau = k/12$, $k = 1, 2,...,12$, are generally not directly available. Hence, an uniform rainfall intensity within each hour is assumed to compute the cumulative rainfall fractions–an equivalence to a linear interpolation of dimensionless rainfall mass curves. The assumption of uniform intensity within each hour may not distort the basic shape of a rainfall pattern when the rainfall duration is relatively long. Without demanding a fine detail on rainfall pattern for establishing a design rainfall hyetograph, rainfall events with duration of at least 3 h might be adequate to reflect the overall temporal variation of hyetograph. Hence, hourly based rainfall events with duration less than or equal to 2 h are excluded from the database. In total, 8289

Fig. 1 Map showing the locations of raingauges used
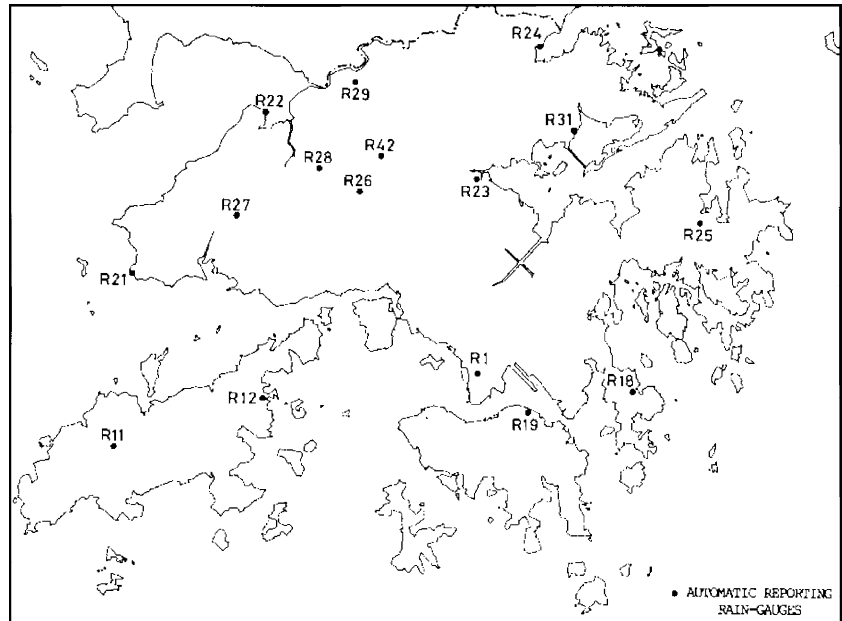
rainfall events are extracted from 16 automatic rainga-uges to identify representative hourly based rainfall patterns in Hong Kong Territory. The contributions of each station to the total number of rainfall events in the database are given in the last column of Table 1.

Cluster analysis for identification representative rainfall patterns

The principle component analysis and cluster analysis have been used to classify hyetographs and hydrographs (Zurich 1971; Fang and Tung 1996; Hannah et al. 2000; Lana et al. 2001; Lin et al. 2004). In addition, Michaelides et al. (2001) used artificial neural network approach to classify 1-year and 2-year rainfall events. The basic aim of cluster analysis is to find the 'natural grouping' of a set of individuals (Collins 1990) characterized by a list of relevant attributes. In other words, cluster analysis allocates a set of individuals to a set of mutually exclusive and exhaustive groups such that individuals within the group are similar to one another whereas individuals between different groups are dissimilar. To remove the scale effects of the attributes used in the cluster analysis, attributes are standardized resulting in a zero mean and a unit standard deviation. In general, the divisive (K-means method) and hierarchical (average and Ward's methods) clustering techniques can be applied to perform the cluster analysis and each method defines the similarity and the distance measurement somewhat differently. The results from different applications (Fang and Tung 1996; Ramos2001) revealed that the effects of

Table 1 List of raingauge stations used in the study

Station no.	Location	Elevation (m)	Region	Record period used	No. of rainfall events used
R01	Hong Kong Observatory	32	Kowloon	1884–1939, 1947–1996	4359
R11	Ngong Ping Tea Farm	440	Islands	1984–1996	180
R12	Discorvery Bay Water Treatment Works	75	Islands	1984–1996	297
R18	Sam Yuk Middle School	105	Kowloon	1985–1996	284
R19	Quarry Bay Tide Gauge House	10	HK Island	1992–1996	132
R21	Tap Shep Kok Power Station	25	NT West	1984–1996	328
R22	Tsim Bei Tsui Meteo. Station	5	NT West	1984–1996	222
R23	Wong Shiu Chi Middle School	25	NT East	1984–1996	417
R24	Sha Tau Kok Police Station	35	NT East	1984–1996	324
R25	Pak Tam Au Country Park	105	Kowloon	1984–1996	305
R26	Shek Kong RAF Airfield	10	NT West	1985–1996	332
R27	Yuen Long R.G. Filters	90	NT West	1985–1996	211
R28	Au Tau Fish Farm	5	NT West	1985–1996	215
R29	Lok Ma Chau Police Station	50	NT West	1985–1996	241
R31	Tai Mei Tuk Pumping Station	10	NT East	1985–1996	358
R42	Fire Dept. Training School, Yuen Long	10	NT West	1992–1996	84

using methods of K-means, Ward, and others on the final classification result were insignificant. Moreover, Ramos (2001) also indicated that K-means method seems to be powerful enough to classify the observation. Hence, the Euclidean distance-based K-means clustering method (MacQueen1967) is adopted herein to classify 8289 rainfall events into a few representative rainfall patterns typically occurred in Hong Kong Territory. The K-means clustering algorithm is performed by using a statistical software Minitab (Minitab Inc., 1996).

According to the study on the relative performance of using different attributes to classify rainfall patterns in Wyoming (Fang and Tung 1996), the parametric way of using statistical moments or fitted distribution parameters as the attributes resulting in a less desirable classification of rainfall patterns when compared with a direct use of dimensionless rainfall mass or hyetograph ordinates. Hence, the latter is considered herein by which dimensionless rainfall mass curves ordinates F_1, F_2, ..., F_{11}, denoted as F-based ordinates, and dimensionless rainfall hyetograph ordinates P_1, P_2, ..., P_{12}, denoted as P-based ordinates, are used to identify representative rainfall patterns in Hong Kong.

Since the appropriate number of representative rainfall patterns is not known in advance, it is commonly deter-mined by a trial-and-error process. The proper number of rainfall patterns can be determined by visually examining the averaged dimensionless mass curves for each group resulting from the cluster analysis. In this study, 3–7 groups of rainfall patterns are examined using the ordinates of both dimensionless rainfall mass curve and hye-tograph. Figures 2, 3, and 4 show the averaged rainfall mass curves under 5, 6, and 7 groups of classification. It is clearly observed that the 6-group classification produced a distinct uniform-like rainfall pattern, not revealed in a 5-group classification. Comparing 6-group and 7-group classification results (Figs. 3, 4), the difference is that the central-peaked pattern in the 6-group classification is further splitted into two patterns with slightly advanced and slightly delayed peaking times. Such a fine distinction of rainfall patterns is judged unnecessary and, therefore, six patterns are considered representative of typical rain-fall temporal distribution profiles occurred in Hong Kong Territory. By comparing the results of F-based and P-based classifications, the results are similar. However, the rainfall patterns obtained by the F-based classification are more desirable, according to a subjective visual judgment, because the vertical spacing between the resulting mass curves is somewhat wider than the P-based classi-cation.

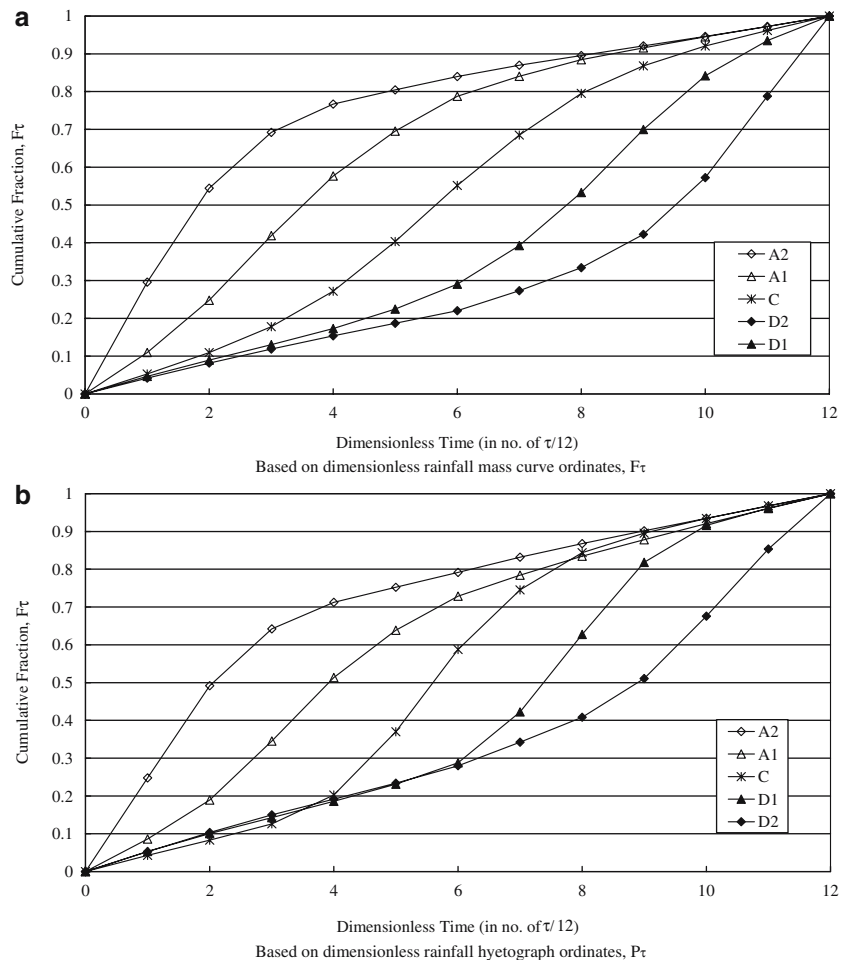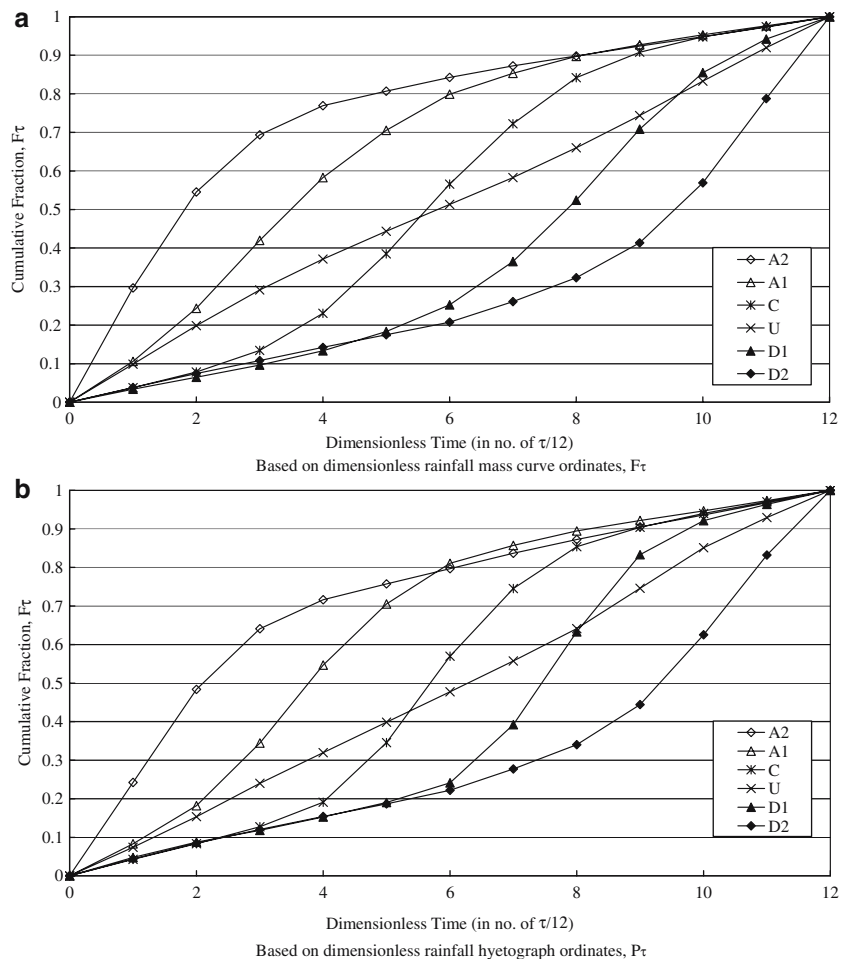**Fig. 2** Classification of 5-group rainstorm patterns

Fig. 3 Classification of 6-group rainstorm patterns

a

Based on dimensionless rainfall mass curve ordinates, F_τ

b

Based on dimensionless rainfall hyetograph ordinates, P_τ

In summary, Fig. 3a is recommended as the representative rainfall patterns (for $T \geq 3$ h) typically occurred in Hong Kong Territory. They consist of four basic types of rainfall patterns: advanced type (A_1 and A_2), central-peaked type (C), delayed type (D_1 and D_2), and uniform type (U). The advanced patterns have relatively high rainfall intensity during early part of the rainfall event, whereas the delayed type is just the opposite. The central-peaked pattern has relatively high rainfall intensity in the center part of the rainfall event and the intensity tapers off towards the beginning and ending of the rainfall event, while the uniform type reveals relatively constant rainfall intensity throughout the rainfall period. The mean values of dimensionless rainfall mass curve and hyetograph ordinates of the six rainfall patterns are given in Table 2.

Factors affecting the occurrence frequency of rainfall patterns

Note that the six representative rainfall patterns shown in Fig. 3a are derived from the pool of dimensionless rainfall mass curves of many rainfall events with different durations and depths that have occurred in different seasons at various locations in Hong Kong Territory. To facilitate proper selection of rainfall patterns in the design and analysis of hydrosystem infrastructures or in stochastic generation of rainfall events, it is important to investigate whether the occurrence frequency of one particular rainfall pattern might be affected by the season, geographical location, or rainfall duration/depth.

To examine the possible inter-dependence between the occurrence frequency of rainfall pattern and geographical location, rainfall depth and depth, and seasonality, statistical contingency table analysis can be performed. The objective of the statistical contingency table analysis is to decide whether the occurrence of various rainfall patterns is affected by or is independent of the different factors under consideration (Conover 1980). The contingency tables for the six representative rainfall patterns in Hong Kong Territory are shown in Table 3a–d. This table shows the occurrence frequency of various rainfall patterns with respect to different levels of the various factors under consideration. In the study, the chi-square test is conducted by the statistical software Minitab to compute the p-value, which indicates the probability that the occurrence of various rainfall patterns is independent of the specific factor under consideration. Information on occurrence frequencies summarized in the contingency tables and the

Table 2 Mean values of dimensionless rainfall mass curve and hyetograph ordinates for the six representative rainfall patterns in Hong Kong Territory

Rainfall pattern[a]	Dimensionless rainfall mass											
	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}
A_2	0.247	0.476	0.646	0.757	0.803	0.843	0.875	0.906	0.934	0.958	0.981	1.000
A_1	0.072	0.180	0.360	0.558	0.708	0.808	0.867	0.911	0.940	0.962	0.982	1.000
C	0.032	0.067	0.115	0.196	0.347	0.539	0.713	0.846	0.912	0.950	0.977	1.000
U	0.123	0.243	0.347	0.424	0.483	0.538	0.602	0.678	0.759	0.846	0.932	1.000
D_1	0.029	0.057	0.086	0.118	0.163	0.226	0.336	0.501	0.704	0.868	0.952	1.000
D_2	0.039	0.077	0.110	0.141	0.172	0.203	0.236	0.278	0.350	0.504	0.756	1.000
Rainfall pattern	Dimensionless rainfall hyetograph											
	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
A_2	0.247	0.229	0.170	0.111	0.046	0.040	0.032	0.031	0.028	0.024	0.023	0.019
A_1	0.072	0.108	0.180	0.197	0.150	0.100	0.058	0.045	0.029	0.022	0.020	0.018
C	0.032	0.035	0.048	0.081	0.151	0.191	0.175	0.132	0.066	0.038	0.027	0.023
U	0.123	0.120	0.103	0.077	0.060	0.055	0.064	0.076	0.081	0.086	0.086	0.068
D_1	0.029	0.028	0.028	0.032	0.045	0.063	0.110	0.165	0.203	0.163	0.085	0.048
D_2	0.039	0.037	0.033	0.031	0.031	0.031	0.034	0.041	0.072	0.154	0.252	0.244

A advanced type, *C* central-peaked type, *U* uniform type, *D* delayed type

results from such analysis can help proper selection of rainfall profiles for various engineering studies.

Three numbers are shown in each cell of Table 3a–d, namely, frequency count, row percentage, and column percentage. At the bottom of each table, it presents the values of chi-square test statistic for the data, the degree of freedom associated with the chi-square test, which equal to $(n_c - 1)(n_r - 1)$ with n_c and n_r being the numbers of column and row, respectively, and the *p*-value. In practice, the significance level of 5% is used for decision-making. If the *p*-value is lower than 5%, the factor under consideration is considered to have significant effect on the occurrence frequency of various rainfall patterns.

Referring to the computed *p*-values shown in the bottom row of Table 3a–d, it can be concluded that the occurrence frequency of various rainfall patterns in Hong Kong Territory are affected by the season, rainfall duration and depth, but not by the geographical location. In all contingency tables, it is clearly shown that the distribution of the occurrence frequency of the six rainfall patterns (indicated by the row percentages) do not vary uniformly from one level to another for the factor under consideration. Table 3a and b shows the general trend that the occurrence frequency for all rainfall patterns, except the uniform type, decreases with rainfall duration and depth. For seasonality effect, Table 3c shows that it is five to six times more likely to have a rainfall event of any pattern in wet season than in dry season. Table 3d shows that there is a much higher occurrence frequency in Kowloon and this is because Hong Kong Observatory (Station R01) contributes more than half of the retrieved rainfall events in the database.

For rainfall duration, depth, and seasonality, Table 3a–c reveal that the distribution of the occurrence frequency across different levels of the three factors (see column percentages) varies from one rainfall pattern to another. On the other hand, Table 3d shows that the distribution of column percentages of the six rainfall patterns stays rather constantly within a geographical region.

The results from the contingency tables analysis imply that the choice of selecting or generating a particular rainfall pattern for engineering application in Hong Kong should not be arbitrary as the occurrence frequency of different rainfall patterns is dependent on the rainfall depth and duration. Information on the occurrence frequency of different rainfall patterns provides the basis for choosing rainfall types that are most likely to occur or for conducting risk-based analysis considering uncertainty in the occurrence of various rainfall patterns.

Stochastic generation of rainfall patterns

In the stochastic generation of realizable rainfall events, once the depth and duration of rainfall event are specified the follow-up task is the generation of time distribution of rainfall intensities, i.e., rainfall hyetograph, conditioned on the rainfall duration and depth. From the results of rainfall pattern classification, it is observed that individual rainfall event of a particular pattern is unique by itself, meaning temporal patterns of individual rainfall events are similar but not identical. The representative rainfall patterns shown in Fig. 3a are the averaged rainfall mass curves for each pattern and there exists intrinsic variations in rainfall shapes within each type. To generate probabilistically plausible rainfall patterns of a certain type, one must recognize two important features: (1) ordinates of the dimensionless rainfall mass curve are bounded, $0 \leq F_\tau \leq 1$, and non-decreasing; (2)

Table 3 Contingency tables showing occurrence frequency of six typical rainfall patterns by different factors in Hong Kong Territory

Rainfall duration (h)	Rainfall patterns						Column total
	A_1	A_2	C	D_1	D_2	U	
3–6	714[a]	760	833	480	345	201	3333
	21.42[b]	22.80	24.99	14.40	10.35	6.03	100.00
	39.96[c]	58.28	40.42	31.33	50.29	21.94	40.22
6–9	403	252	377	324	109	172	1637
	24.62	15.39	23.03	19.79	6.66	10.51	100.00
	22.55	19.33	18.29	21.15	15.89	18.78	19.76
9–12	217	116	250	209	73	107	972
	22.33	11.93	25.72	21.50	7.51	11.01	100.00
	12.14	8.90	12.13	13.64	10.64	11.68	11.73
12–18	257	127	295	245	88	175	1187
	21.65	10.70	24.85	20.64	7.41	14.74	100.00
	14.38	9.74	14.31	15.99	12.83	19.10	14.33
>18	196	49	306	274	71	261	1157
	16.94	4.24	26.45	23.68	6.14	22.56	100.00
	10.97	3.76	14.85	17.89	10.35	28.49	13.96
Row total	1787	1304	2061	1532	686	916	8286
	21.57	15.74	24.87	18.49	8.28	11.05	100.00
	100.00	100.00	100.00	100.00	100.00	100.00	100.00
$\chi^2 = 570.27$	$DF = 20$	p value $= 0.00$					
Rainfall depth (mm)							
0–20	631[a]	757	687	543	395	239	3252
	19.40[b]	23.28	21.13	16.70	12.15	7.35	100.00
	35.31[c]	58.05	33.33	35.44	57.58	26.06	39.24
20–40	555	334	670	458	169	333	2519
	22.03	13.26	26.60	18.18	6.71	13.22	100.00
	31.06	25.61	32.51	29.90	24.64	36.31	30.40
40–80	384	170	425	330	84	216	1609
	23.87	10.57	26.41	20.51	5.22	13.42	100.00
	21.49	13.04	20.62	21.54	12.24	23.56	19.42
80–120	120	30	124	108	17	70	469
	25.59	6.40	26.44	23.03	3.62	14.93	100.00
	6.72	2.30	6.02	7.05	2.48	7.63	5.66
>120	97	13	155	93	21	59	438
	22.15	2.97	35.39	21.23	4.79	13.47	100.00
	5.43	1.00	7.52	6.07	3.06	6.43	5.29
Row total	1787	1304	2061	1532	686	917	8287
	21.56	15.74	24.87	18.49	8.28	11.07	100.00
	100.00	100.00	100.00	100.00	100.00	100.00	100.00
$\chi^2 = 471.09$	$DF = 20$	p value $= 0.00$					
Season							
Dry	266[a]	143	310	259	55	112	1145
	23.23[b]	12.49	27.07	22.62	4.80	9.78	100
	14.89[c]	10.97	15.04	16.91	8.02	12.21	13.82
Wet	1521	1161	1751	1273	631	805	7142
	21.30	16.26	24.52	17.82	8.84	11.27	100
	85.11	89.03	84.96	83.09	91.98	87.79	86.18
Row total	1787	1304	2061	1532	686	917	8287
	21.56	15.74	24.87	18.49	8.28	11.07	100
	100	100	100	100	100	100	100
$\chi^2 = 46.84$	$DF = 5$	p value $= 0$					
Region							
HK Island	25[a]	23	30	25	12	17	132
	18.94[b]	17.42	22.73	18.94	9.09	12.88	100
	1.40[c]	1.76	1.46	1.63	1.75	1.85	1.59
Islands	101	58	110	88	53	68	478
	21.13	12.13	23.01	18.41	11.09	14.23	100
	5.65	4.45	5.34	5.74	7.73	7.42	5.77
Kowloon	1059	786	1266	923	387	527	4948
	21.40	15.89	25.59	18.65	7.82	10.65	100
	59.26	60.28	61.43	60.25	56.41	57.47	59.71
NT East	234	168	263	197	103	130	1095
	21.37	15.34	24.02	17.99	9.41	11.87	100
	13.09	12.88	12.76	12.86	15.01	14.18	13.21

Table 3 (Contd.)

Rainfall duration (h)	Rainfall patterns						Column total
	A_1	A_2	C	D_1	D_2	U	
NT West	368	269	392	299	131	175	1634
	22.52	16.46	23.99	18.30	8.02	10.71	100
	20.59	20.63	19.02	19.52	19.10	19.08	19.72
Row total	1787	1304	2061	1532	686	917	8287
	21.56	15.74	24.87	18.49	8.28	11.07	100
	100	100	100	100	100	100	100
$\chi^2 = 23.20$	$DF = 20$	p value $= 0.279$					

[a] Frequency counts
[b] Row percentage
[c] Column percentage

ordinates of the dimensionless rainfall mass curve or hyetograph are possibly correlated and most likely they do not have normal distributions.

Hence, the problem of synthesizing rainfall patterns in essencoe is the one to generate multivariate non-normal random variables subject to the following two constraints:

$$\text{unit - sum}: P_1 + P_2 + \cdots + P_K = 1.0 \tag{1a}$$

$$\text{Non - negativity}: P_\tau \geqslant 0, \tau = 1, 2, \ldots, K \tag{1b}$$

in which K is the number of ordinates defining the dimensionless rainfall hyetograph ($K = 12$ in this study). The random variables in vector $\mathbf{P} = (P_1 + P_2 + \ldots + P_K)^t$ satisfying the above constraint Eqs. 1a and b are called compositional variables. Aitchison (1986) presents many examples and problems related to compositional data analysis in different disciplines. The

Fig. 4 Classification of 7-group rainstorm patterns
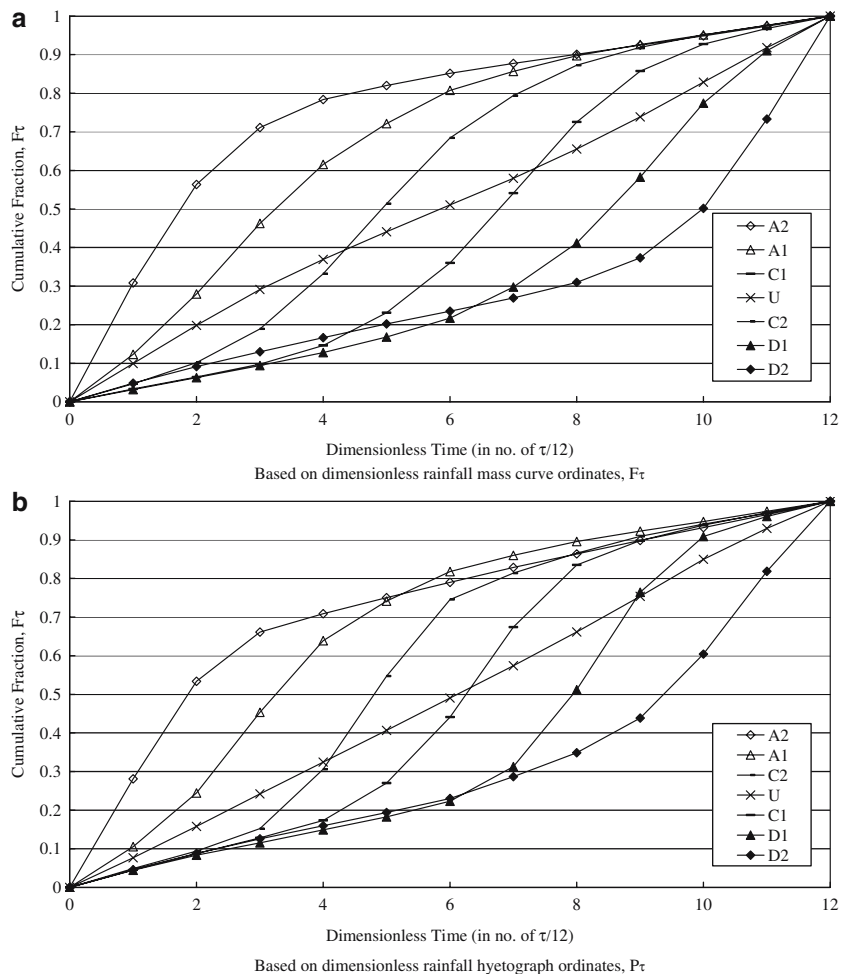
generation of compositional variable with prescribed statistical properties, such as statistical moments and correlations, is not a trivial task due to the presence of the constraints. As the constrained Monte–Carlo simulation involves generation of unconstrained, non-normal random variables, the following two subsections provide a brief overview of the procedures for normal transform and stochastic generation of unconstrained as well as constrained multivariate non-normal random variates.

Generation of non-negative multivariate random variates under unit-sum constraint

Procedures for constrained multivariate normal Monte–Carlo simulation are available. Borgman and Faucette (1993) presented a practical method to convert a linearly constrained multivariate Gaussian simulation into a conditional multivariate Gaussian simulation (Borgman1990). Zhao (1992) applied the method to generate random unit hydrographs and to evaluate the reliability of hydraulic structures due to uncertainty in unit hydrograph (Zhao et al. 1997). Lambert and Kuczera (1996) alternatively proposed a procedure by which ordinates of dimensionless rainfall hyetographs, P_τ, are treated as conditionally truncated log-normal random variables.

To generalize the procedure for simulating non-negative, non-normal, correlated random variables subject to unit-sum constraint, Fang and Tung (1996) examined three procedures, namely, the acceptance–rejection method, the CDF method, and the log-ratio method. It was found that the log-ratio method is flexible and computational robust than the other two. For this reason, the log-ratio method is adopted in this study to generate dimensionless rainfall hyetographs.

The method is developed by Aitchison (1986), which involves the log-ratio transform as

$$Y_\tau = \log(P_\tau/P_{\tau*}) \quad \text{for } \tau = 1, 2,...,K; \quad \tau \neq \tau* \tag{2}$$

where P_τ can be any element in the compositional variable vector. Since $0 \leq P_\tau \leq 1$ for $\tau = 1, 2,..., K$, the transformed variable Y_τ would range from $-\infty$ to ∞. Note that in the above transform, the sample values of neither P_τ nor P_τ can be 0 to avoid numerical problem. The inverse transform of the log-ratio method yields the following logistic functional relationship between Y_τ and P_τ,

$$P_\tau = \frac{\exp(Y_\tau)}{1+\sum_{\substack{r=1 \\ r\neq\tau*}}^{K}\exp(Y_r)}, \quad 1 \leq \tau \leq K; \tau \neq \tau*$$

$$P_{\tau*} = \frac{1}{1+\sum_{\substack{r=1 \\ r\neq\tau*}}^{K}\exp(Y_r)} \tag{3}$$

It is obvious that $0 \leq P_\tau \leq 1$ and $P_1 + P_2 + ... + P_K = 1$.

After performing the log-ratio transform, the problem of generating constrained multivariate non-normal variates P_τs is converted to the problem of generating unconstrained multivariate non-normal random variates Y_τs. In Aitchison (1986), a multivariate normal distribution is assumed for Y_τs, which results in a logistic-normal distribution. However, the normality condition for the log-ratio, Y_τ, is not satisfied in this study. Therefore, the Johnson distribution system is selected to describe the marginal distribution of random variables in vector $Y = (Y_1, Y_2,..., Y_{\tau*-1}, Y_{\tau*+1},..., Y_K)^t$.

The computational steps of the log-ratio method for generating dimensionless rainfall hyetographs are as follows:

Step 1. Choose τ^* and the corresponding $P_{\tau*}$ to carry out the log-ratio transform on compositional data by Eq. 2.
Step 2. Calculate sample moments and correlations of random log-ratio Y_τ, $\tau = 1, 2,..., \tau^* - 1, \tau^* + 1,..., K$.
Step 3. Fit Y_τ individually to the Johnson distribution system for all $\tau \neq \tau*$.
Step 4. Generate Y_τ for all $\tau \neq \tau *$ using the procedures for simulating unconstrained multivariate non-normal random variates described in Sect. 6.2.
Step 5. Obtain dimensionless rainfall hyetograph ordinates $P_1, P_2,..., P_K$ by Eq. 3.

Generation of unconstrained multivariate non-normal random variates

In dealing with problems involving multivariate non-normal random variables, the main difficulty is the availability or derivation of the joint probability density function. Based on many efficient procedures developed for generating multivariate normal random variates, the common approach to simulate unconstrained multivariate non-normal random variates involves three basic steps: (1) transformation of non-normal random variables into the normal space; (2) generation of multivariate normal random variates; (3) inverse transformation back to the original space. Tadikamalla (1980) provided an good review on transforming univariate non-normal random variables to the normal space in step 1. A practical way to transform multivariate non-normal-to-normal can be made through predefined marginal distributions and correlation structure (Liu and Der Kiureghian1986; Chang et al. 1997). As marginal distributions of random variables are often unknown in practice, normal transform for a non-normal random variable is made parametrically by assuming an appropriate distribution or non-parametrically on the basis of data or derived statistical properties (Chen and Tung 2003). In this study, the parametric approach using the Johnson distribution system is adopted for its flexibility of covering a wide variety of distribution types.

Johnson ([1949](#)) introduced a system of frequency functions consisting of four parameters:

$$Z = g(X|\gamma, \delta, \xi, \lambda) = \gamma + \delta f\left(\frac{X-\xi}{\lambda}\right) \tag{4}$$

where Z is a standard normal variable, X the original non-normal variable, γ, δ, ξ, λ are model parameters. There are three special types of the Johnson distribution:

$$S_L : Z = \gamma + \delta\ln(X-\xi), \quad X > \xi \tag{5}$$

$$S_U : Z = \gamma + \delta\sinh^{-1}\left(\frac{X-\xi}{\lambda}\right), \quad X > \xi \tag{6}$$

$$S_B : Z = \gamma + \delta\ln\left(\frac{X-\xi}{\xi+\lambda-X}\right), \quad \xi < X < \xi+\lambda \tag{7}$$

in which S_L is a log-normal distribution, S_U the unbounded distribution, S_B is a bounded distribution. Hill et al. ([1976](#)) provided an algorithm to estimate model parameters by matching the first four product-moments of X and attempted to identify one of the best-fitted Johnson distribution type.

To obtain the correlation structure in the normal space as required in step 2 of normal transform, the Johnson distribution system can be extended to the multivariate setting using the Nataf bivariate model (Nataf [1962](#)) as

$$\begin{aligned}\rho_{x_i,x_j} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{x_i-\mu_i}{\sigma_i}\right)\left(\frac{x_j-\mu_j}{\sigma_j}\right)\phi_2(z_i,z_j;\rho_{z_i,z_j})\mathrm{d}z_i\mathrm{d}z_j(z_j)\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{g_{x_i}^{-1}(z_i)-\mu_i}{\sigma_i}\right)\left(\frac{g_{x_j}^{-1}(z_j)-\mu_j}{\sigma_j}\right)\\ &\quad \times \phi_2(z_i,z_j;\rho_{z_i,z_j})\mathrm{d}z_i\mathrm{d}z_j\end{aligned} \tag{8}$$

where ρ_{x_i,x_j} and ρ_{z_i,z_j} are correlation coefficients in the original and normal spaces, respectively, μ_i and σ_i the mean and standard deviation of the random variable X_i, $g_{x_i}^{-1}(z_i)$ the inverse function of the Johnson frequency function, $\phi_2(z_i, z_j)$ is a bivariate standard normal probability density function. Empirical formulas similar to those developed by Liu and Der Kiureghian ([1986](#)) have been derived which relate the correlation coefficient in the normal space ρ_{z_i,z_j} to that in the original parameter space ρ_{x_i,x_j}, the distribution parameters, and the first two moments of some combinations of the Johnson distribution types (Fang and Tung [1996](#)). For the combinations of more complicated distribution types, one has to resort to Eq. 8 for each individual evaluation.

As a practical alternative to Eq. 8, correlation coefficients in the normal space $\mathbf{R}_z = \left[\rho_{z_i,z_j}\right]$ can be estimated by transforming sample observations in the original space $\{x_i\}$ to the standard normal space $\{z_i\}$ through the use of Eq. 4 with the distribution parameters estimated from the sample moments. Upon the availability of the correlation structure of involved random variables in the normal space, $\mathbf{R}_z$, random variates of unconstrained multivariate non-normal random variables can be generated by the following steps:

Step 1. Based on the covariance or correlation matrix in an m-dimensional multivariate normal space, determine an $m \times m$ orthogonal transformation matrix $\mathbf{T}$ that decomposes the correlated standard normal random variables $\mathbf{Z}$ into uncorrelated ones $\mathbf{Z}'$ as $\mathbf{Z}' = \mathbf{T}^{-1}\mathbf{Z}$. Such an orthogonal transformation matrix can be obtained by the well-known Chelosky factorization or spectral decomposition (Tung and Yen [2005](#)).

Step 2. Generate m independent standard normal random variates $\mathbf{z}' = (z'_1, z'_2, \dots z'_m)^t$ by an appropriate algorithm, from which the corresponding correlated standard normal variates can be obtained as $\mathbf{z} = \mathbf{T}\mathbf{z}'$.

Step 3. Multivariate non-normal random variates can be computed by the inverse transform through $x_i = F_i^{-1}[\Phi(z_i)]$ for $i = 1, 2, \dots, m$ with $F_i^{-1}(x)$ being the inverse CDF of the original random variable X_i; $\Phi(.)$ is the standard normal CDF.

Numerical example

Based on the results of classifying more than 8000 rainfall events in Hong Kong Territory shown above, the statistical features of dimensionless rainfall hyetograph ordinates for each pattern can be summarized. For example, Table 4 lists the first four moments of the dimensionless rainfall hyetograph ordinates and their correlations for rainfall patterns of advanced type A_2 based on 1787 individual rainfall events in the category. Figure 5 shows the mean and 95% confidence bands envelope curves for advanced type A_2.

To generate statistically realizable rainfalls of this pattern by the log-ratio method, Eq. 2 is applied, after selecting a base ordinate $P_{\tau*}$, to compute the log-ratio values $Y_\tau = \ln(P_\tau/P_{\tau*})$ for $\tau \neq \tau$ of each rainfall event within the group. By taking $\tau^* = 3$, the sample statistics of Y_τ based on rainfall events containing nonzero hyetograph ordinates among the 1787 type-A_2 cases are computed (see upper part of Table 5).

For a given rainfall pattern, from the first four sample statistical moments of log-ratios Y_τ, the best fit Johnson distribution type and the corresponding parameters for each Y_τ are determined (see lower part of Table 5 for type A_2 rainfall as an example). According to the distribution type and parameters for each Y_τ, sample values (z's) in the standard normal space corresponding to sample log-ratio values (y's) are computed via Eqs. 5, 6, 7 based on which the correlation matrix among normally transformed log-ratio random variables is calculated. By following the procedure described in Sect. 6.2, a sequence of the correlated standard normal random vector z_τ can be obtained from which the corresponding log-ratio value can be

Table 4 Summary statistics of dimensionless rainfall hyetograph ordinates for advanced rainfall pattern A$_2$

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
Mean	0.247	0.229	0.170	0.111	0.046	0.040	0.032	0.031	0.028	0.024	0.023	0.019
SD	0.071	0.070	0.079	0.078	0.039	0.032	0.028	0.024	0.030	0.020	0.023	0.016
Skewness	1.146	1.272	−0.380	0.196	2.079	2.055	3.350	1.874	3.911	1.818	2.606	2.453
Kurtosis	5.958	8.389	2.665	1.442	9.213	8.568	22.078	8.292	24.740	6.813	11.585	11.734
P_1	1.000											
P_2	−0.195	1.000										
P_3	−0.457	0.160	1.000									
P_4	−0.484	−0.320	0.261	1.000								
P_5	−0.105	−0.165	−0.356	−0.042	1.000							
P_6	0.032	−0.273	−0.359	−0.224	0.479	1.000						
P_7	0.059	−0.251	−0.382	−0.214	0.020	0.294	1.000					
P_8	0.046	−0.239	−0.329	−0.190	−0.010	0.034	0.671	1.000				
P_9	0.195	−0.296	−0.373	−0.323	−0.030	0.118	0.236	0.439	1.000			
P_{10}	0.143	−0.173	−0.468	−0.258	0.022	0.134	0.071	0.041	0.404	1.000		
P_{11}	0.208	−0.066	−0.390	−0.220	−0.087	−0.054	−0.041	−0.103	0.075	0.536	1.000	
P_{12}	0.190	−0.128	−0.291	−0.161	−0.133	−0.145	−0.067	−0.022	0.113	0.415	0.677	1.000

computed as $y_\tau = g_\tau^{-1}(z_\tau | \gamma, \delta, \xi, \lambda)$ in which $g_\tau^{-1}(z_\tau | \gamma, \delta, \xi, \lambda)$ is the inverse Johnson distribution function associated with the random log-ratio variable Y_τ. With the vectors of random log-ratio variates available, Eq. 3 can be applied to compute the ordinates of dimensionless rainfall hyetographs. For example, according to the statistical features of the dimensionless rainfall hyetograph ordinates for rainfall pattern A$_2$ (Table 4) and those of the corresponding log-ratio variates (Table 5), a sample of 15 plausible rainfall mass curves of pattern-A$_2$ are shown in Fig. 6.
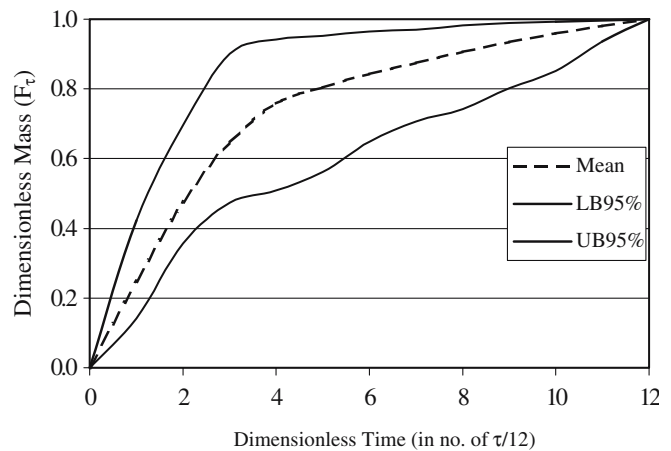

Fig. 5 Mean and 95% confidence bands of advanced rainstorm pattern-A2

Summary and conclusions

The temporal rainfall pattern often is a required input to rainfall–runoff modeling for producing flow hydro-

Table 5 Summary statistics of log-ratios, P_τ / P_3, for advanced rainfall pattern A_2

	Y_1	Y_2	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}	Y_{11}	Y_{12}
Mean	0.536	0.449	−0.611	−1.409	−1.540	−1.764	−1.765	−1.966	−2.092	−2.191	−2.289
SD	0.962	0.806	1.100	1.297	1.296	1.302	1.231	1.378	1.362	1.314	1.233
Skewness	1.507	2.036	−1.102	0.553	0.697	0.827	0.797	0.821	0.607	0.634	0.690
Kurtosis	5.325	8.068	4.106	3.322	4.634	5.057	5.159	4.332	3.287	3.198	3.578
Y_1	1.000										
Y_2	0.817	1.000									
Y_4	0.335	0.281	1.000								
Y_5	0.722	0.683	0.500	1.000							
Y_6	0.748	0.630	0.381	0.859	1.000						
Y_7	0.744	0.657	0.352	0.692	0.778	1.000					
Y_8	0.729	0.635	0.357	0.673	0.686	0.873	1.000				
Y_9	0.770	0.637	0.242	0.611	0.639	0.726	0.771	1.000			
Y_{10}	0.793	0.675	0.345	0.637	0.654	0.685	0.661	0.851	1.000		
Y_{11}	0.747	0.673	0.314	0.569	0.547	0.585	0.558	0.712	0.864	1.000	
Y_{12}	0.745	0.660	0.344	0.577	0.541	0.574	0.580	0.745	0.850	0.915	1.000
Distribution	S_U S_U	S_U	S_U	S_B	S_B	S_B	S_B	S_U	S_U	S_U	
γ	1.727	2.114	−1.712	3.135	−1.196	−1.278	−1.057	−4.237	2.207	1.688	3.261
δ	0.886	0.846	1.135	2.618	2.357	2.236	2.126	3.331	2.034	1.672	2.342
λ	6.588	6.816	8.041	19.006	2.437	2.218	2.048	2.243	14.775	11.450	17.843
ξ	−0.583	−0.345	−6.941	−5.945	−2.951	−3.243	−2.950	−5.822	−5.978	−5.416	−5.990

Note: $Y_\tau = \ln(P_\tau/P_z3)$ for $\tau \neq 3$

Fig. 6 A sample of simulated advanced rainstorm pattern-A2
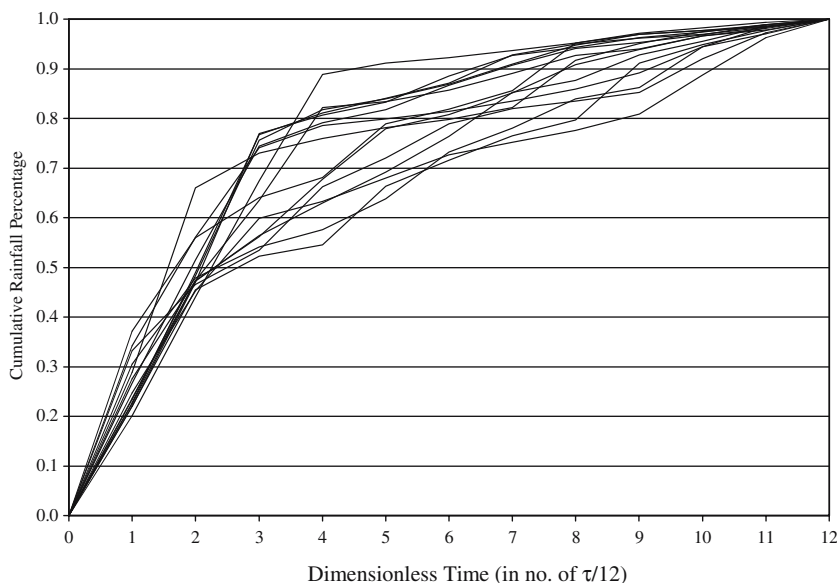
graphs in design and analysis of hydrosystems. In this study, event-based representative temporal rainfall patterns throughout Hong Kong Territory have been identified using statistical cluster analysis. Through non-dimensionalization, ordinates of rainfall mass curves and hyetographs are used as the attributes in statistical cluster analysis. In this study, it is judged that the use of dimensionless mass curve results in a better classification of six rainfall patterns.

In this study, statistical contingency table analysis is applied to determine whether the occurrence frequency of the various rainfall patterns in Hong Kong Territory are affected by factors such as the rainfall duration, depth, seasonality, and geographical location. It was found that the occurrence frequency of the six representative rainfall patterns in Hong Kong Territory are affected by the rainfall duration, depth, and seasonality, but not affected by the geographical location. Information such as this is useful for proper selection of rainfall patterns in the stochastic generation of rainfall events and rainfall–runoff modeling.

Since the rainfall pattern is one of important rainfall characteristics often required in hydrosystem engineering study, this paper also presents a practical procedure for generating probabilistically plausible rainfall hyetographs of a particular pattern. The procedure involves the log-ratio transform of compositional variables and normal transform of non-normal multivariate random variables, which can be straightforwardly implemented.

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 & Hall Inc., NY, USA

Bandyopadhyay M (1972) Synthetic rainfall pattern and run-off for Gauhati, India. J Hydraul Div ASCE 98(HY5):845–857

Borgman LE (1990) Irregular ocean waves: kinematics and forces. In: Mehaute BL, Hanes DM (eds) Ocean engineering science. Wiley Inc., NY, USA

Borgman LE, Faucette RC (1993) Frequency-domain simulation and stochastic interpolation of random vectors in multidimensional space. In: Cheng HD, Yang CY (eds) Computational stochastic mechanics. Elsevier, London, UK

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

Chen CL (1976) Urban rainfall runoff inlet hydrograph study, vol 4. Synthetic rainfalls for design of urban highway drainage facilities. Report FHWA-RD-76-119. Federal Highway Administration, U.S. Department of Transportation, Washington, DC, USA

Chen XY, Tung YK (2003). Investigation of polynomial normal transformation. J Struct Saf 25:423–445

Collins AJ (1990) Introduction to multivariate analysis. Chapman & Hall Inc., London

Conover WJ (1980) Practical nonparametric statistics. Wiley, NY, USA

Fang TQ, Tung YK (1996) Analysis of wyoming extreme precipitation patterns and their uncertainty for safety evaluation of hydraulic structures. Technical Report, WWRC-96-05, Wyoming Water Resources Center, University of Wyoming, Laramie, Wyoming

Hannah DM, Smith BPG, Gurnell AM, McGregor RM (2000) An approach to hydrograph classification. Hydrol Process 14:317–338

Heneker TM, Lambert MF, Kuczera G (2001) A point rainfall model for risk-based design. J Hydrol 247:54–71

Hershfield DM (1962) Extreme rainfall relationships. J Hydraul Div ASCE 88(HY6):73–92

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

Huff FA (1967) Time distribution of rainfall in heavy rainfalls. Water Resour Res 3(4):1007–1019

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

Keifer CJ, Chu HH (1957) Synthetic rainfall pattern for drainage design. ASCE J Hydraulic Div 83(HY4):1–25

Lambert M, Kuczera G (1996) A stochastic model of rainfall and temporal pattern. In: Tickle K et al (eds) Stochastic hydraulics '96. AA Bulkema Publishers, Rotterdam, The Netherlands, pp 317–324

Lana X, Serra C, Burgueno A (2001) Patterns of monthly rainfall shortage and excess in terms of the standardized precipitation index for Catalonia (Ne Sapin). Int J Climatol 21:1669–1691

Lin GF, Chen LH, Kao SC (2004) Development of regional design hyetographs. Hydrol Process (in press)

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

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

Michaelides SC, Pattichis CS, Kleovoulou G (2001) Classification of rainfall variability by using artificial neural networks. Int J Clinatol 21:1401–1414

Nataf A (1962) Determination des distributions de probabilit_s dont les marges sont donn_es. Comput Rendus Acad Sci Paris 255:42–43

Ogrosky HL (1964) Hydrology of spillway design: Small structure-limited data. ASCE J Hydraulic Div 90(HY3):42–43

Onof C, Wheater HS (1993) Modelling of Britsh rainfall using a Barlett-Lewis rectangular pulse model. J Hydrol 149:67–95

Pilgrim DH, Cordery I (1975) Rainfall temporal patterns for design floods. J Hydraulic Div ASCE 101(HY1):81–95

Ramos MC (2001) Divisive and hierarchical clustering techniques to analyse variability of rainfall distribution patterns in a Mediterranean region. Atmos Res 57:123–138

Soil Conservation Service (SCS) (1972) National engineering handbook, Sect. 4: hydrology. US Department of Agriculture, Washington, DC, USA

Stuart A, Ord JK (1987) Kendall's advanced theory of statistics, vol 1—distribution theory, 5th edn. Oxford University Press, NY, USA

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

Tung YK, Yen BC (2005) Hydrosystem engineering uncertainty analysis, McGraw-Hill Book Company, NY, USA

US Army Corps of Engineers (USACOE) (1948) Hydrological and hydraulic analyses, flood hydrograph analysis and compilations, Chap. 5. Engineering Manual Civil Works Construction, Part CXIV

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

Yen BC, Chow VT (1983) Local design rainfall, vol II—methodology and analysis. Report, FHWA/RD-82/064, Federal Highway Administration, US Department of Transportation, Washington, DC, USA

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

Zhao B (1992) Determination of a unit hydrograph and its uncertainty applied to reliability analysis of hydraulic structures. MS Thesis, Department of Statistics, University of Wyoming, p 352

Zhao B, Tung YK, Yeh KC, Yang JC (1997) Reliability analysis of hydraulic structures considering unit hydrograph uncertainty. J Stoch Hydrol Hydraulic 11(1):33–50

Zurich AK (1971) Statistical-analytical methods in theoretical regional design hyetograph. Geoforum 7:39–53