Weather and Climate Extremes 43 (2024) 100634

Contents lists available at ScienceDirect

Weather and Climate Extremes

journal homepage: www.elsevier.com/locate/wace

Rainfall frequency analysis using event-maximum rainfalls – An event-based mixture distribution modeling approach

Ke-Sheng Cheng [a,b,*], Bo-Yu Chen [b], Teng-Wei Lin [a], Kimihito Nakamura [c], Piyatida Ruangrassamee [d], Hidetaka Chikamori [e]

[a] *Department of Bioenvironmental Systems Engineering, National Taiwan University, Taipei, Taiwan, ROC*
[b] *Master Program in Statistics, National Taiwan University, Taipei, Taiwan, ROC*
[c] *Graduate School of Agriculture, Kyoto University, Kyoto, Japan*
[d] *Department of Water Resources Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand*
[e] *Department of Environmental Management Engineering, Okayama University, Okayama, Japan*

ARTICLE INFO

Keywords:
Frequency analysis
Event-maximum rainfalls
Mixture distribution
Annual maxima
Poisson distribution
Negative binomial distribution

ABSTRACT

Rainfall frequency analysis, an essential work for water resources management, is often conducted by using the annual maximum rainfall series. For rainfall stations with short record lengths and outliers presence, the use of annual maximum series for rainfall frequency analysis may yield design rainfall estimates of higher uncertainties. Moreover, for regions with cyclostationary climate patterns, the annual maximum rainfalls may be caused by different prevalent storm types, which differ in terms of their occurrence frequency and storm rainfall characteristics. In this study, we propose a novel event-maximum-rainfall-based mixture distribution modeling approach for rainfall frequency analysis. By considering the event-maximum rainfalls of individual storm events, the sample size for parameter estimation increases, and the uncertainty of design rainfall estimates reduces. Mixture distribution modeling enables a thorough investigation of the contributing probabilities of different storm types to the annual maximum rainfall. Through rigorous stochastic simulation, we demonstrated the superiority of the proposed approach over the conventional annual maximum rainfall approach. The proposed approach was applied to four representative rainfall stations in Taiwan, and the results revealed that the proposed approach is more robust than the conventional annual maximum rainfall approach. The results provide insights into the contributions of individual storm types to the annual maximum rainfall.

1. Introduction

Rainfall frequency analysis, which aims to determine the magnitude of rainfall corresponding to a given recurrence interval, is essential for designing hydraulic structures and flood risk management. The conventional approach of at-site rainfall frequency analysis establishes the intensity-duration-frequency (IDF) curves based on the annual maximum rainfall (AMR) of specific design durations. In Taiwan, design durations of 1, 2, 3, 6, 12, 24, 48, and 72 h and return periods of 5, 10, 25, 50, 100, and 200 years are often used to construct IDF curves. To determine the design rainfall (i.e., the rainfall magnitude for the desired return periods), probability distributions must be fit to the AMR of specific design durations. Goodness-of-fit tests such as the chi-squared test, Kolmogorov-Smirnov test, and L-moment-ratio-diagram-based test together with the information criteria are applied to identify the best-fit distribution (Cheng et al., 2011).

Design rainfall represents the magnitude of low probability of exceedance, and it is sensitive to the shape parameter of the selected probability distribution. Therefore, parameter estimation of the best-fit distribution plays a crucial role in rainfall frequency analysis. The AMR series of a specific design duration observed at a rainfall station forms a random sample for parameter estimation. Several factors affect the accuracy of parameter estimation, including the sample size, presence of outliers, and the parameter estimation method. A short record length yields an estimate with high uncertainty. In addition, outliers, which are extraordinary values with a very low likelihood of occurrence, markedly affect the accuracy of parameter estimation, and, consequently, design rainfall estimates. For instance, in August 2009, Typhoon Morakot caused heavy rainfall in southern Taiwan, and many rainfall stations documented record-breaking amounts of rainfall in 24, 48, and 72 h (see

* Corresponding author. Department of Bioenvironmental Systems Engineering, National Taiwan University, Taipei, Taiwan, ROC.
E-mail address: rslab@ntu.edu.tw (K.-S. Cheng).

https://doi.org/10.1016/j.wace.2023.100634
Received 11 August 2023; Received in revised form 30 November 2023; Accepted 30 November 2023
Available online 8 December 2023

Table 1
Observed rainfall depths at 15 rainfall stations in the Kaoping River watershed in southern Taiwan during the overpass of Typhoon Morakot. (Adapted from WRA, 2009).

Station	*t*-hour event-maximum rainfall depth (*d*, in mm) and its estimated return period (*T*, in years)						Event total depth (mm)	Record length (years)
	24-h		48-h		72-h			
	d	T	d	T	d	T		
Ping-Tung (5)	667	141	886	124	947	159	959	38
Mei-Nong (2)	507	>2000	749	>2000	828	>2000	871	19
Ping-Tung	666	140	906	143	975	197	990	38
Xi-Pu	730	271	995	265	1058	378	1077	38
Qi-Shan	621	>2000	813	>2000	855	>2000	881	15
Weiliao-Shan	1415	>2000	2216	>2000	2564	>2000	2701	21
Jia-Sien	1078	>2000	1601	>2000	1856	>2000	1916	25
Gu-Xia	684	>2000	946	>2000	1062	>2000	1127	25
Mei-Nong	634	>2000	878	>2000	956	>2000	990	15
Li-Gun	711	>2000	956	>2000	1018	>2000	1040	15
Shanderwen	1186	>2000	1968	>2000	2195	>2000	2255	25
Xin-Wei	578	148	758	>2000	807	565	831	25
Yue-Mei	744	>2000	1081	>2000	1205	>2000	1247	19
Ji-Dong	548	>2000	728	>2000	789	>2000	821	19
Dah-Jin	739	>2000	1072	>2000	1241	>2000	1314	21

Table 1). The IDF relationships established using historical records (excluding the observed rainfall caused by Typhoon Morakot) revealed that many of the recorded rainfall amounts exceeded the 2000-year return level (WRA, 2009). Similarly, an extraordinary flood occurred in Lockyer Valley, Queensland, Australia, in 2011. Using the annual maximum flood series up to 2010 (a record length of 31 years), the return period of the 2011 flood was estimated to be approximately 2000 years (Rogencamp and Barton, 2012; Thompson and Croke, 2013). When 3 more years of flood data (including the data for the 2011-2013 floods) were incorporated into the annual flood series, a return period of 59 years was estimated for the 2011 flood event (Sargood et al., 2015).

Although the annual maxima are widely used in rainfall frequency analysis, when only the maximum value in each year is considered, other valuable information in the collected rainfall data may be neglected (Volpi et al., 2019; Lee et al., 2020). To overcome this disadvantage, the peak-over-threshold (POT) method is recommended for rainfall frequency analysis. The POT method requires additional computational efforts to select only those peaks that exceed the threshold (i.e., the maximum of a cluster of values all exceeding the threshold) and additional information or parameters to select independent peak values (Coles, 2001).

For the POT method, a threshold value is first determined. Then all peak rainfall amounts exceeding the threshold value are obtained and fit to the Pareto distribution for deriving the design rainfall depths. Selection of the threshold value is a subjective process and is disadvantageous for the POT method (Bezak et al., 2014). The threshold value should be high enough for the threshold exceedances to be mutually independent. Moreover, A minimum time interval between two peaks is required to define two separate events. However, the threshold value selected should be as low as possible to yield a large number of exceedances and more reliable parameter estimates (Ashkar and Rousselle, 1987). There are various criteria for the threshold selection, including the lowest annual maximum value (Langbein, 1949), which is equivalent to the frequency factor in the range 3–3.5 (Madsen et al., 1993; Madsen and Rosbjerg, 1997), and a level which, on average, four events per year were selected (Bačová-Mitková and Onderka, 2010). By setting a

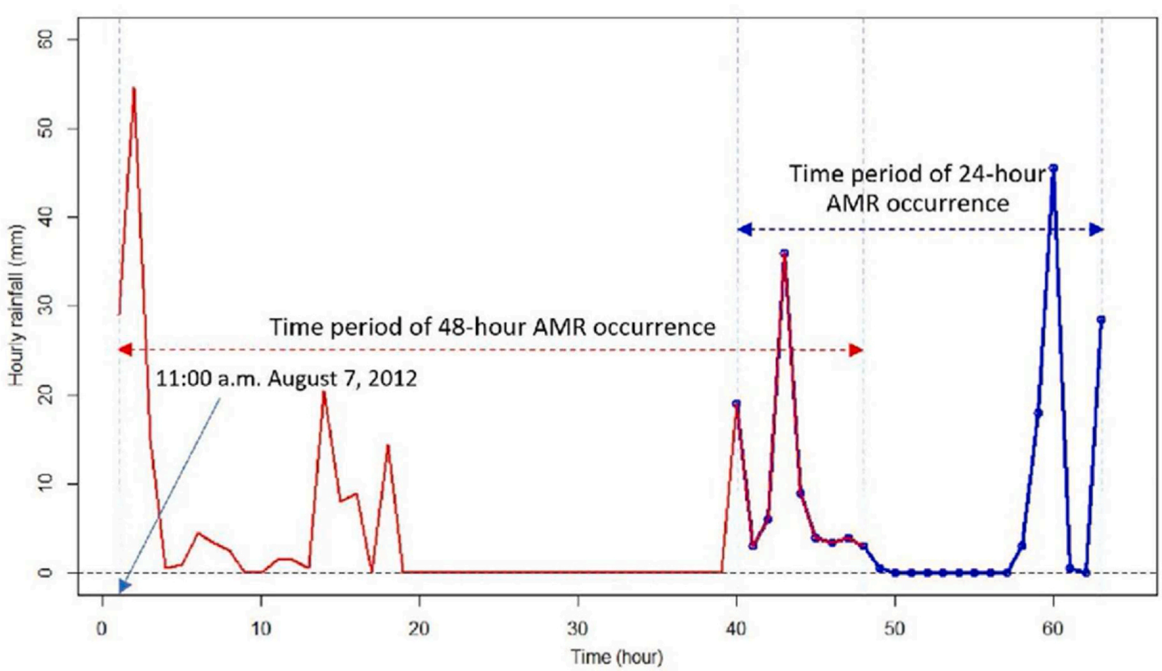

Fig. 1. The 24 and 48-h annual maximum rainfall (183.5 and 253.5 mm, respectively) at the Kaoshiung station in 2012.

minimum time interval between two peaks, the POT method collects data from independent events.

Automated raingauge networks (ARNs) have been established in many countries in recent years. In Taiwan, the Central Weather Agency (CWA) established an ARN of more than 500 rainfall stations. Till now, most of these stations have provided less than 30 years of rainfall observations. Similarly, the Thai Meteorological Department, the Royal Irrigation Department, the Department of Water Resources, and the Hydro-Informatics Institute (HII) maintain automatic weather stations and the hydromet telemetry system in Thailand. The HII has established more than 700 ARN stations that have recorded rainfall data since 2007.

East and Southeast Asia frequently experience tropical monsoons and typhoons. Extreme rainfalls caused by these events are often considered outliers at ARN stations with short records, and it is disadvantageous to calculate the design rainfalls using the AMR series. Therefore, this study proposed a new approach of rainfall frequency analysis by using the event-maximum rainfall (EMR), and the performance of the proposed approach was compared with that of the conventional AMR approach through stochastic simulation.

2. Background for extreme-value data series and modeling

2.1. Data series for hydrological frequency analysis

For rainfall frequency analysis, random samples of extreme rainfalls are needed. These samples can be collected based on the concept of the annual maximum series (AMS), annual exceedance series (AES), or POT series. The AMS and AES have sample sizes equal to the record length (in years) of the observed rainfall data, whereas the sample size of a POT series is usually larger than the record length. The quality of extreme quantile estimation using the POT series mainly depends on the choice of the threshold. A high threshold yields high-variance estimators, and a low threshold may result in more biased estimators (El-Aroui and Diebolt, 2002). For short-record rainfall stations, frequency analysis using AMS or AES data yields design rainfall estimates with higher uncertainty. The POT series increases the sample size, and a few studies have demonstrated that it can provide better estimates of design rainfall depths [Madsen et al., 1997; Claps and Liao, 2003; Bezak et al., 2014]. However, no clear criterion is available for determining the threshold for extreme rainfall extraction.

During the extraction of the AMS, AES, or POT data from historical rainfall measurements, the following concerns should be considered:

(1) Traditionally, the AMR of the *t*-hour design duration is determined by continuously moving a window of *t*-hour width along the 1-year timeline. The AMR of longer design durations (for example, 24 or 48 h) may comprise rainfall of different storm events separated by a dry period of several hours. As demonstrated in Fig. 1, the 24-h and 48-h AMR at the Kaoshiung station in 2012 were caused by two storm events separated by 8-h and 21-h dry periods, respectively. Most of the watersheds in Taiwan have time of concentration of only a few hours. Storm events separated by an interval larger than the time of concentration of the drainage basin do not contribute jointly to the streamflow at the basin outlet. Therefore, the AMR caused by multiple events separated by long dry periods should not be included in the annual maximum series. Hereinafter, we refer to the AMR extracted using a *t*-hour moving window as the *traditional* AMR.

(2) A minimum time interval separating POT data should be considered to ensure that data in the POT series are mutually independent (Rustomji et al., 2009). In general, successive storm events should be separated by sufficiently long periods with no rainfall. Naghettini (2017) suggested a dry period interval of 1–2 days for daily rainfall and a dry period interval of 6 h for subdaily rainfall. Similar considerations should be applied to flood frequency analysis by using the POT series. A minimum period, during which discharge must be below the threshold value, is required for consecutive floods to be considered independent. For example, a threshold of 15 days was used for data collected from three large catchments in Australia (Karim et al., 2017).

The aforementioned concerns necessitate the use of an event-based hydrological frequency analysis method. The first concern is related to the extreme rainfall generated by a single event, whereas the second concern focuses on the extreme values of independent events. Rainfall and floods are caused by individual storm events; thus, storm characteristics, including the occurrence frequency, duration, and total rainfall depth of storm events, play a crucial role in the analysis of rainfall and flood frequencies.

2.2. Block maximum and the extremal type theorem

Probability distributions such as log-normal, Pearson type III (PE3), log-Pearson type III, Gumbel, generalized extreme value (GEV), and generalized Pareto (GP) distributions have been used to model the AMR or the POT rainfall. The best-fit distribution is usually selected based on results of the goodness-of-fit test (D'Agostino and Stephens 1986; Hosking and Wallis 1997; Liou et al., 2008) and based on the information criteria (Akaike, 1974; Schwarz, 1978; Cheng et al., 2011). The US Water Resources Council has recommended the log-Pearson type III distribution for flood modeling in the United States (Stedinger et al., 1993). The PE3 and log PE3 distributions are widely used for rainfall frequency analysis in Taiwan. In Europe, many studies have used the GEV distribution for rainfall frequency analysis (Deidda et al., 2021).

Gumbel (1954, 1963) developed the theory of extreme values and proposed three types of extreme value distributions – the Gumble, Fréchet, and Weibull distributions. The extremal type theorem (ETT) unifies the three types of extreme value distributions.

Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed (iid) continuous random variables with a common cumulative distribution function (CDF) F_X. The maximum of $X_1, X_2, \cdots, X_n$ is termed the *block maximum* of block size *n*. Let $M_n = max(X_1, X_2, \cdots, X_n)$, then the CDF of M_n is:

$$P[M_n \leq x] = [F_X(x)]^n \tag{1}$$

The block maxima, when properly normalized, can only converge to one of the three types of extreme value distributions, Gumbel, Fréchet, and Weibull. The GEV distribution is a family of continuous probability distributions that combines the Gumbel, Fréchet, and Weibull distributions. By the ETT, the GEV distribution is the only possible non-degenerative limiting distribution of the normalized block maxima of a sequence of iid random variables. However, in implementing this model, the choice of block size is critical. Blocks that are too small are likely to yield a poor approximation by the GEV distribution, leading to bias in estimation and extrapolation; large blocks generate few block maxima, leading to large estimation variance. Pragmatic considerations often lead to the adoption of blocks of length one year (Coles, 2001).

The speed of convergence of M_n to the limiting distribution is dependent on the underlying distribution F_X. For example, convergence is slow for maxima of *n* Gaussian variables (Huxham and Mcgilchrist, 1969; Coles and Davison, 2008). Thus, for block maxima of small to moderate block size, whether the GEV fits the available data needs to be empirically assessed.

3. Mixture distribution modeling of the annual maximum rainfalls

3.1. Cyclostationary climatological process

Hydrological and climatological data are often described by time series models. A stationary time series process denotes that the probability distribution of hydrological or climatological variables does not

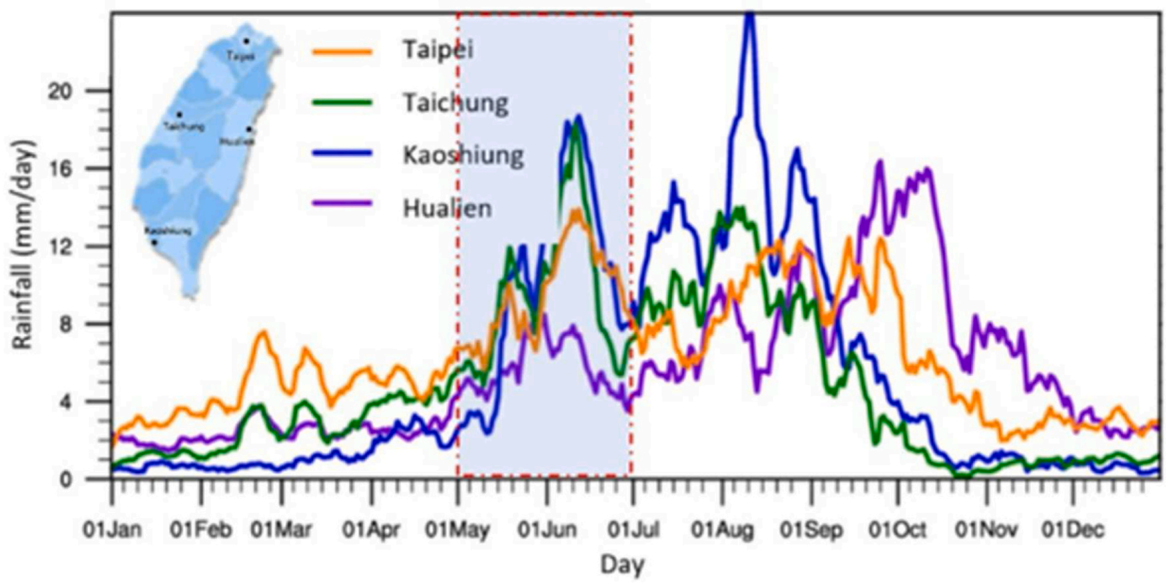

Fig. 2. The cyclostationary rainfall pattern in Taiwan. The 9-day running average rainfalls are calculated using the 30-year (1991-2020) daily rainfall data. The shaded region represents the Meiyu season. (Adapted from CWA, 2023).

change with time. A cyclostationary process denotes that the probability distribution changes periodically with time; that is, the first and second moments depend on an external annual or diurnal cycle (von Storch et al., 1995). Many geophysical and climatological processes such as precipitation and global average surface temperature may be termed cyclostationary (Kim, 2000).

In regions with complex climate regimes, various prevalent storms can occur in different seasons. For instance, southern China, Taiwan, and Japan experience heavy rainfalls caused by mesoscale convective systems (MCSs), which is known as the Meiyu frontal rainfalls, in late spring or early summer, and by typhoons in summer or early fall. In Taiwan, seasonal prevalent storms, including Meiyu, summertime convective storms, typhoons, and winter-to-spring (WS) frontal rainfall, occur in different months. These prevalent storms differ in terms of their occurrence frequency, storm duration, and rainfall intensity.

Meiyu rainfall is produced by surface frontal systems that advance northeasterly from southern China to Taiwan from May to the end of June, each year. Generally, 3–5 Meiyu frontal systems are observed in a year. During the passage of a Meiyu frontal system, a few very active mesoscale convective cells may repeatedly develop, causing heavy rainfall in the area. Although the synoptic-scale frontal system may last for a few days, the MCSs generally have lifetimes ranging from a few hours to less than 1 day (Cheng et al., 2008).

Thunderstorms and showers associated with afternoon convection are common weather phenomena in summer over Taiwan and often lead to flooding locally. In July and August, an average of approximately 13 days of convective storm events is observed (Chen and Chou, 2006). Most of these convective storms have short durations of only a few hours.

Tropical cyclones are the most severe meteorological threat to Taiwan. Typically, 21–25 typhoons form in the northwestern Pacific from June to December, and 3–5 typhoons make landfalls in Taiwan annually. Although tropical cyclones can last for several days to a week, local precipitation durations are typically 18–36 h in Taiwan.

WS (November to April) frontal rainfall is caused by the northeasterly monsoon, particularly over the northeastern part of Taiwan. On average, 39 rainy days of WS frontal rainfall are observed in a year (Chang et al., 2019). The WS frontal events have long durations (i.e., typically longer than 1 day); however, their rainfall intensities are considerably lower than those of other types of storms.

Because summertime convective storms only last a few hours (mostly <6 h) and WS frontal events have lower rainfall intensities than do other storm types, the annual maxima of long design durations (e.g., 24 and 48 h) are mostly produced by typhoons and Meiyu storms. However, short-duration annual maxima can be formed by storms of either type. The likelihood of the AMR caused by a particular storm type depends on the storm characteristics, including the occurrence frequency, duration, and amount and time variation of event-total rainfall, of individual storm types.

The occurrences of seasonal prevalent storms result in a nonstationary annual rainfall process. Fig. 2 presents the 9-day running average of daily rainfalls for four representative rainfall stations in Taiwan (CWA, 2023). The intrayear nonstationarity is due to different characteristics of seasonal prevalent storms. Under the assumption that the characteristics of the prevalent storms remain unchanged, the rainfall process can be considered to be cyclostationary with an annual cycle.

The ETTdescribed in Section 2.2 requires all M_n values to be of the same block size and be identically distributed. For rainfall frequency analysis under the cyclostationary climatological condition, annual maximum rainfalls may be contributed by different types of prevalent storms which differ in their rainfall characteristics. Additionally, the annual number of prevalent storms varies across years; thus, the AMR may comprise block maxima of varying block sizes. Therefore, under the cyclostationary climatological condition, the annual maxima do not satisfy the assumptions of the ETT. Allamanno et al. (2011), Strupczewski et al. (2012), and Mascaro (2018) have studied the effects of nonstationarity (seasonality) on the distribution of hydrological extremes and frequency analysis. In this study, we propose the following mixture distribution modeling approach for rainfall frequency analysis under the cyclostationary climatological condition.

3.2. Mixture distribution modeling

Suppose that N years of hourly rainfall observations are available at one site, and that the rainfall depth of the t-hour design duration and T-year return period is to be estimated. Let n represent the annual number of typhoons that can be characterized by a Poisson distribution of the following probability mass function,

$$p(n) = \frac{e^{-\lambda}\lambda^n}{n!}, n = 0, 1, 2, \cdots \tag{2}$$

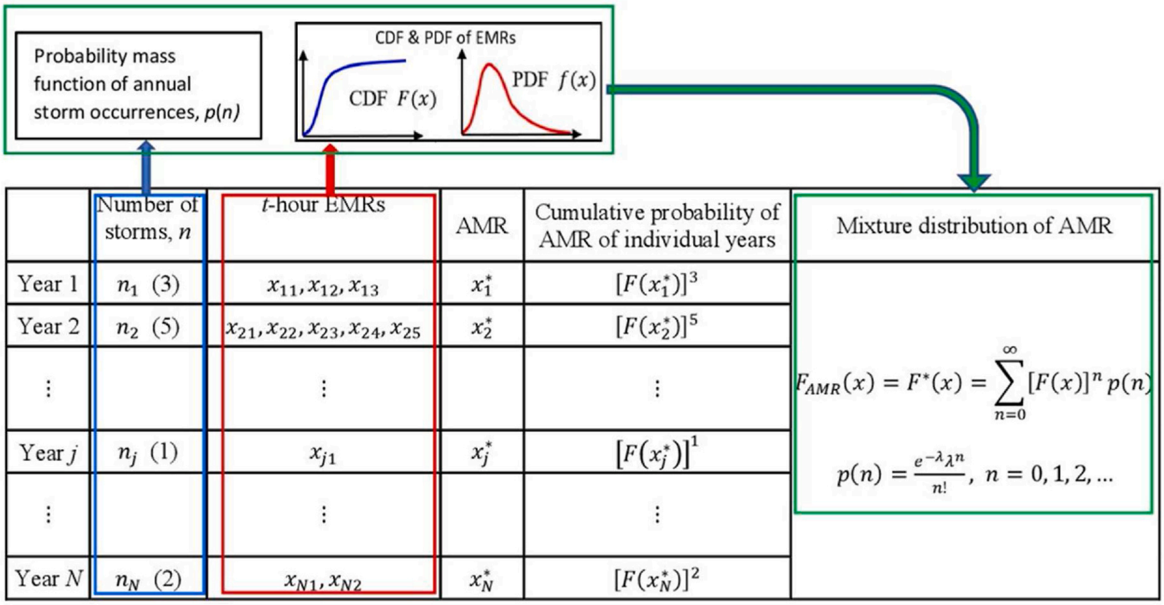

Fig. 3. Storm-type-specific relationship between the EMR and AMR probability distributions.

$$E(n) = Var(n) = \lambda \tag{3}$$

Let X represent the t-hour EMR of a typhoon event. For a year of n typhoon events, the t-hour AMR caused by typhoons is the maximum of the n EMRs, and its probability distribution can be expressed as

$$F_{AMR}(x|n) = [F_X(x)]^n \tag{4}$$

where F_{AMR} and F_X represent the CDFs of the AMR and EMR, respectively. The annual number of typhoons varies yearly and is governed by the Poisson distribution. Therefore, the probability distribution of the t-hour AMR of typhoons can be expressed using the following mixture distribution:

$$F_{AMR}(x) = F^*(x) = \sum_{n=0}^{+\infty} [F_X(x)]^n p(n). \tag{5}$$

where $p(n)$ is the Poisson mass function (Eq. (2)). Fig. 3 illustrates the conceptual derivation of the above theoretical relationship between the probability distribution function of the EMR and AMR. The probability mass function of the annual number of typhoons and the CDF of the t-hour EMR can be estimated from the observed rainfall data of individual storm events. By considering the event-maximum series (EMS), we increase the sample size for EMR parameter estimation and reduce the uncertainty of parameter estimates. Notably, in contrast to the traditional AMR, the AMR described in this section is event-based with no occurrence of the cross-event total rainfall problem described in Fig. 1.

In the above derivation, only typhoon occurrences and the EMR of typhoons are considered; thus F_{AMR} in Eq. (5) only characterizes the probability distribution of the AMR produced by typhoons. The probability distribution of AMR of a particular prevalent storm type is hereafter referred to as the *storm-type-specific* AMR distribution. For regions under cyclostationary climatological conditions, different storm types may produce the AMR. In our study, four prevalent storm types, namely Meiyu, summertime convective storms, typhoons, and WS frontal events, are considered, and the overall AMR is the maximum of four storm-type-specific annual maxima in the same year. We further assume that AMR distributions of individual storm types are mutually independent. Therefore, the CDF of the overall AMR can be expressed as

$$\begin{aligned} G(x) &= P(X_{all} \le x) = P\left(X_{Mei} \le x, X_{Con} \le x, X_{Typ} \le x, X_{Frn} \le x\right) \\ &= P(X_{Mei} \le x) P(X_{Con} \le x) P\left(X_{Typ} \le x\right) P(X_{Frn} \le x) \\ &= F^*_{Mei}(x) F^*_{Con}(x) F^*_{Typ}(x) F^*_{Frn}(x) \end{aligned} \tag{6}$$

where X_{all}, X_{Mei}, X_{Con}, X_{Typ}, and X_{Frn} represent the random variables of the overall and four storm-type-specific AMR. G and F^* are the CDFs of the overall and storm-type-specific AMR, respectively.

3.3. Evaluating the mixture distribution modeling approach by stochastic simulation

To evaluate the performance of the proposed mixture distribution modeling approach for the AMR, we conducted the following stochastic simulation by considering two prevalent storm types, namely Meiyu storms and typhoons. Hereinafter, the terms "*EMS-based approach*" and "*AMS-based approach*" are used to denote the proposed mixture distribution modeling approach and the conventional annual-maximum-rainfall approach, respectively.

We assume that the 24-h AMR at a rainfall station may be produced by typhoons or Meiyu storms. Let random variables X_1 and X_2 represent the 24-h EMR of typhoons and Meiyu storms, respectively. Both random variables can be characterized by the following gamma density:

Table 2
Theoretical depth, bias, and root-mean-squared-error (RMSE) of the 24-h design rainfalls based on the simulation scenario.

Return period (years)	5	10	25	50	100	200
Theoretical depth (mm)	418	523	658	758	858	957
EMS-based approach						
Bias (mm)	2.30	4.49	6.17	8.01	9.28	11.05
RMSE (mm)	52.76	71.10	96.31	115.95	135.84	155.15
AMS-based approach						
Bias (mm)	3.28	3.24	−1.86	−7.70	−15.97	−25.09
RMSE (mm)	58.43	84.75	130.46	169.94	212.21	256.45

Note: The bias and RMSE were calculated based on 10,000 block simulation runs. Each block simulation run consists of 20 years of annual maximum and event-maximum rainfalls.

Setting the storm parameters (including gamma parameters (α, β) and Poisson parameter λ) for the 24-hour event-maximum rainfall of typhoons and Mei-Yu storms **[$\boldsymbol{\Theta_{Typ}}$ and $\boldsymbol{\Theta_{Mei}}$]**

Calculate the theoretical values of the 24-hour design rainfall depths of the 5, 10, 25, 50, 100, and 200-year return periods **[$\boldsymbol{X_{True}}$]** using Eqs. (8) and (9)

M=10,000 *# Number of block simulation runs*

N=20 *# Number of years in each block simulation run*

for (i in 1:M) {

for (j in 1:N) {

simulate the annual number of typhoon events, n_1

generate the 24-hour event-maximum rainfall of individual typhoon events

simulate the annual number of Mei-Yu storms, n_2

generate the 24-hour event-maximum rainfalls of individual Mei-Yu storms

determine the 24-hour annual maximum rainfall of the j-th year

}

Estimate the storm parameters (α, β, λ) of typhoons and Mei-Yu storms using N years of simulated data. **[$\boldsymbol{\widehat{\Theta}_{Typ}}$ and $\boldsymbol{\widehat{\Theta}_{Mei}}$, block-run-specific]**

Estimate the parameters of the annual maximum rainfalls using N years of annual maximum rainfalls **[$\boldsymbol{\widehat{\Theta}_{AMR}}$, block-run-specific]**

Calculate estimates of the 24-hour design rainfall depths of 5, 10, 25, 50, 100, and 200-year return period by the EMS-based approach **[$\boldsymbol{\widehat{X}_{EMS}}$, block-run-specific]**

Calculate estimates of the 24-hour design rainfall depths of 5, 10, 25, 50, 100, and 200-year return period by the AMS-based approach **[$\boldsymbol{\widehat{X}_{AMS}}$, block-run-specific]**

}

Calculate the bias and RMSE of $\boldsymbol{\widehat{X}_{EMS}}$ with respect to different return periods

Calculate the bias and RMSE of $\boldsymbol{\widehat{X}_{AMS}}$ with respect to different return periods

Fig. 4. Detailed procedure of a demonstrative stochastic simulation of the EMS-based and AMS-based approaches for rainfall frequency analysis.

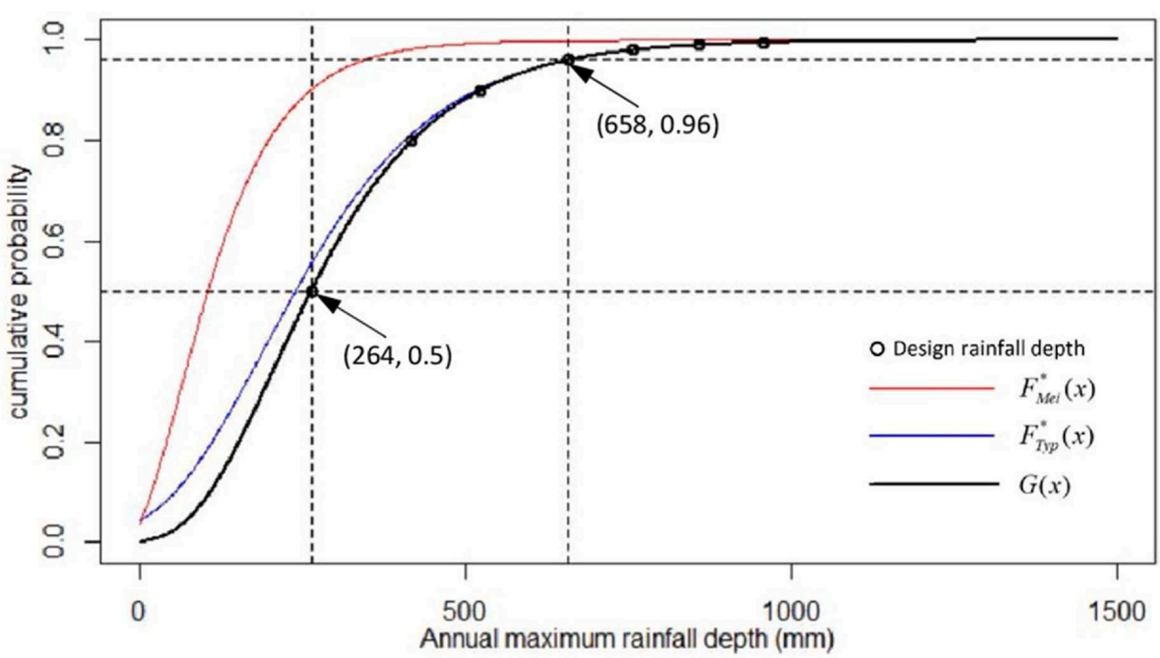

Fig. 5. Theoretical cumulative distribution functions of the overall annual maximum rainfall, $G(x)$, and storm-type-specific annual maximum rainfalls, $F^*_{Mei}(x)$ and $F^*_{Typ}(x)$, based on the simulation scenario described in Section 3.3. Open circle dots represent the 24-h design rainfall depths of the 2, 5, 10, 25, 50, 100, and 200-year return periods.

Table 3
Overall and T-year AMR contributing probabilities of typhoons and Meiyu storms – the simulation scenario.

	Overall	Return period, T, in years						
		2	5	10	25	50	100	200
Typhoons	0.751	0.863	0.924	0.951	0.972	0.983	0.988	0.992
Meiyu	0.249	0.137	0.076	0.049	0.028	0.017	0.012	0.008

$$f(x;\alpha,\beta)=\frac{1}{\alpha\Gamma(\beta)}\left(\frac{x}{\alpha}\right)^{\beta-1}e^{-x/\alpha}, \alpha>0, \beta>0, x\geq 0. \tag{7}$$

$$X_1\sim Gamma(\alpha=141.19, \beta=1.14), X_2\sim Gamma(\alpha=93.87, \beta=0.68) \tag{8}$$

Accordingly, the (mean, standard deviation) pair of the 24-h EMR of typhoons and Meiyu storms are (160.96 mm, 150.75 mm) and (63.83 mm, 77.41 mm), respectively. Let n_1 and n_2 respectively represent the annual number of typhoons and Meiyu storms, which is characterized by the following Poisson distribution:

$$n_1\sim Poisson(\lambda_1=3.09), n_2\sim Poisson(\lambda_2=3.45) \tag{9}$$

Based on the aforementioned parameter settings, the theoretical values of the overall AMR of 5, 10, 25, 50, 100, and 200-year return periods can be calculated using Eqs. (5) and (6) and are shown in Table 2.

We simulated 20 years of typhoon and Meiyu occurrences and the 24-h EMR of typhoons and Meiyu storms according to the number of storms in each year. Based on the simulated data, the scale and shape parameters of X_1 and X_2 were estimated using the method of L-moments (see Appendix A). Parameters n_1, and n_2 were estimated by the mean of the annual number of typhoons and Meiyu storms, respectively. Finally, the CDFs $F^*_{Mei}(x)$ and $F^*_{Typ}(x)$ were estimated, and the probability distribution of the overall AMR, $G(x)$, was calculated to determine the 24-h design rainfall depths of 5, 10, 25, 50, 100, and 200-year return periods. These values are termed the EMS-based design rainfall depths. From the simulated data, the annual maximum was determined as the maximum of all EMR in the same year, including the EMR of typhoons and Meiyu events. The 24-h AMR series was then fitted to a Pearson type III distribution, and the AMS-based design rainfall depths were determined by using the fitted distribution.

The above simulations yielded only one sample estimate of $G(x)$ and one set of estimated design rainfall depths. To compare the performance between the EMS-based and AMS-based approaches, we repeated the above simulation procedure 10,000 times. Each repeat is called a *block simulation run*. Based on the results of 10,000 block simulation runs, the bias and root-mean-squared-error (RMSE) of the EMS-based and AMS-based design rainfall depths were evaluated. The simulation procedures are detailed in Fig. 4.

Fig. 5 presents the theoretical CDFs of the overall and storm-type-specific AMR, which were derived based on the distribution parameter settings for X_1, X_2, n_1, and n_2. The CDF of the overall AMR (i.e., $G(x)$) is nearly identical to that of typhoons, $F^*_{Typ}(x)$, for the AMR higher than or equal to the 25-year return level. This finding indicates that given the simulation scenario, the AMR is mostly caused by typhoons.

3.3.1. AMR contributing probabilities of individual storm types – simulation scenario

The probability that AMR are caused by individual storm types is called the AMR contributing probability, and it can be derived as follows.

Let random variables Y_1 and Y_2 represent the AMR of typhoons and Meiyu storms, respectively. The probability that the AMR is caused by typhoons is equivalent to the probability that Y_1 is higher than Y_2, and it can be expressed by

$$P(Y_1>Y_2)=\int_0^{+\infty}\left[\int_v^{+\infty}f_{Y_1}(u)du\right]f_{Y_2}(v)dv$$

$$=\int_0^{+\infty}[1-F_{Y_1}(v)]f_{Y_2}(v)dv \tag{10}$$

$$\approx\sum_{i=1,2,\ldots}[1-F_{Y_1}(v_i)]f_{Y_2}(v_i)\Delta v$$

Similarly, the contributing probability of typhoons to the T-year AMR, v_T, can be calculated as

$$P(Y_1>\max(Y_2,v_T))=\int_0^{+\infty}\left[\int_{\max(v,v_T)}^{+\infty}f_{Y_1}(u)du\right]f_{Y_2}(v)dv \tag{11}$$

The AMR contributing probabilities calculated using Eqs (10) and (11) are referred to as the overall and the T-year AMR contributing probabilities, respectively.

However, the AMR contributing probabilities cannot be calculated analytically because the probability density function f_{Y_2} in Eqs (10) and (11) cannot be expressed in an analytical form. To overcome this problem, we used the inverse probability integral transformation method based on $F^*_{Mei}(x)$ and $F^*_{Typ}(x)$ to generate 1,000,000 random sample pairs of (Y_1, Y_2). The overall AMR contributing probability of typhoons was then estimated as the proportion of these sample pairs with $Y_1 > Y_2$. The T-year AMR contributing probability of typhoons was calculated in a similar manner, except that only sample pairs with $\max(Y_1, Y_2) > v_T$ were considered. Table 3 presents the AMR contributing probabilities of typhoons and Meiyu storms under the simulation scenario. Typhoons and Meiyu storms contribute approximately 75% and 25% to the AMR, respectively. The T-year AMR contributing probability of typhoons increases with the return period. Approximately 97% of the AMR that are higher than or equal to the 25-year return level are contributed by typhoons.

3.3.2. Bias and RMSE of the EMS-based and AMS-based design rainfall estimates

Based on the results of 10,000 block simulation runs, the frequency histograms and CDFs of the T-year design rainfall estimation errors are presented in Figs. 6 and 7, respectively. The estimation errors in the EMS-based and AMS-based approaches exhibit a normal distribution, although those in the AMS-based approach seem to be slightly right-skewed for higher return periods. The bias and RMSE of the design rainfall estimates of the EMS-based and AMS-based approaches are listed in Table 2. The EMS-based estimates are nearly unbiased, with the highest bias of 1.15% (11.05/957). By contrast, the AMS-based estimates have slightly higher biases, with the highest bias of approximately −2.62% (− 25.09/957). In addition, the EMS-based estimates have lower RMSE values than the AMS-based estimates for all return periods. By using the EMR of individual storm events, the EMS-based approach outperformed the conventional AMS-based approach. Unlike the AMS and the POT series, the EMS is considered to be a complete series because it contains the EMR of all individual storm events, including all the prevalent storm types.

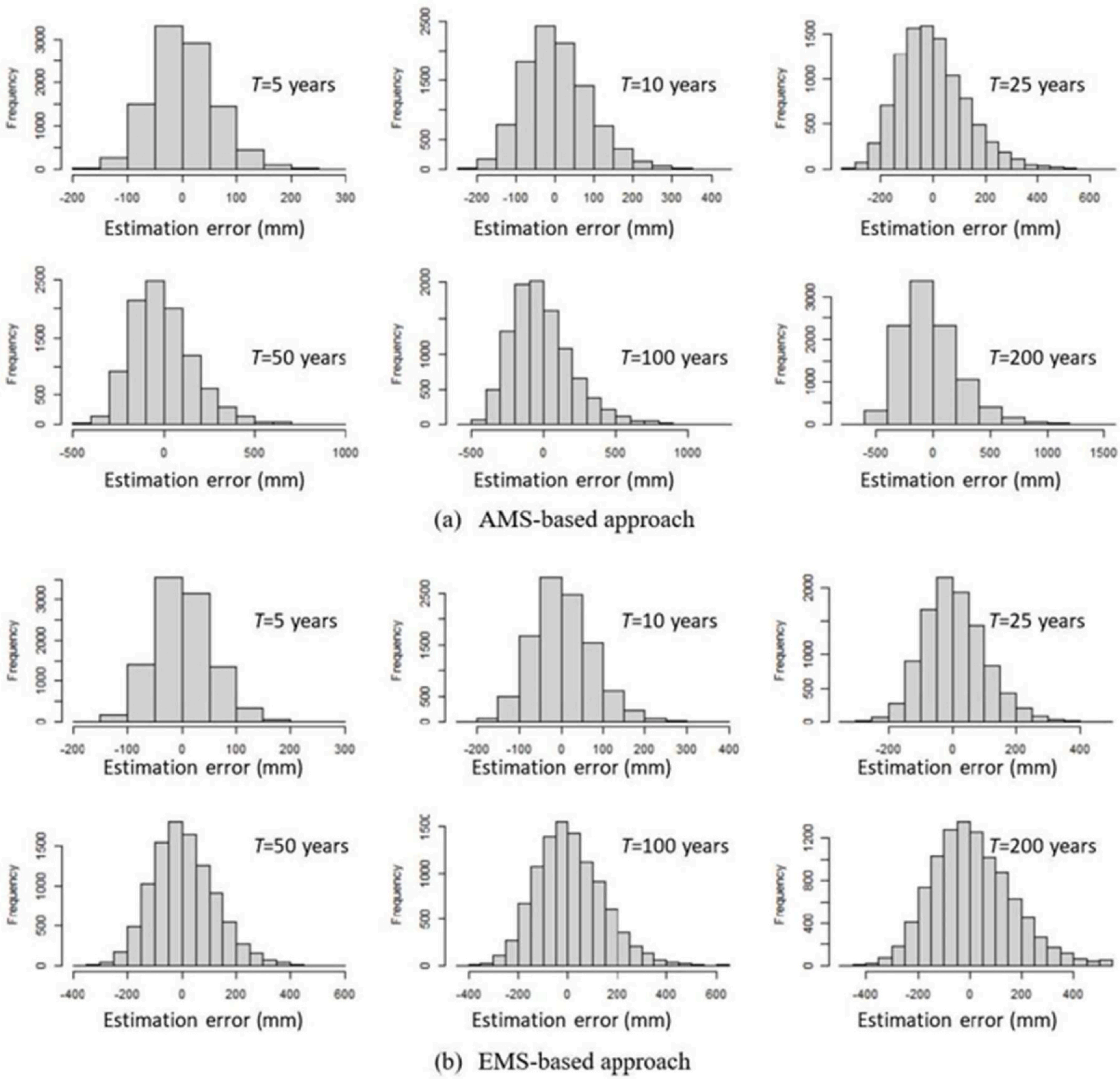

Fig. 6. Histograms of the T-year design rainfall estimation errors based on 10,000 block simulation runs.

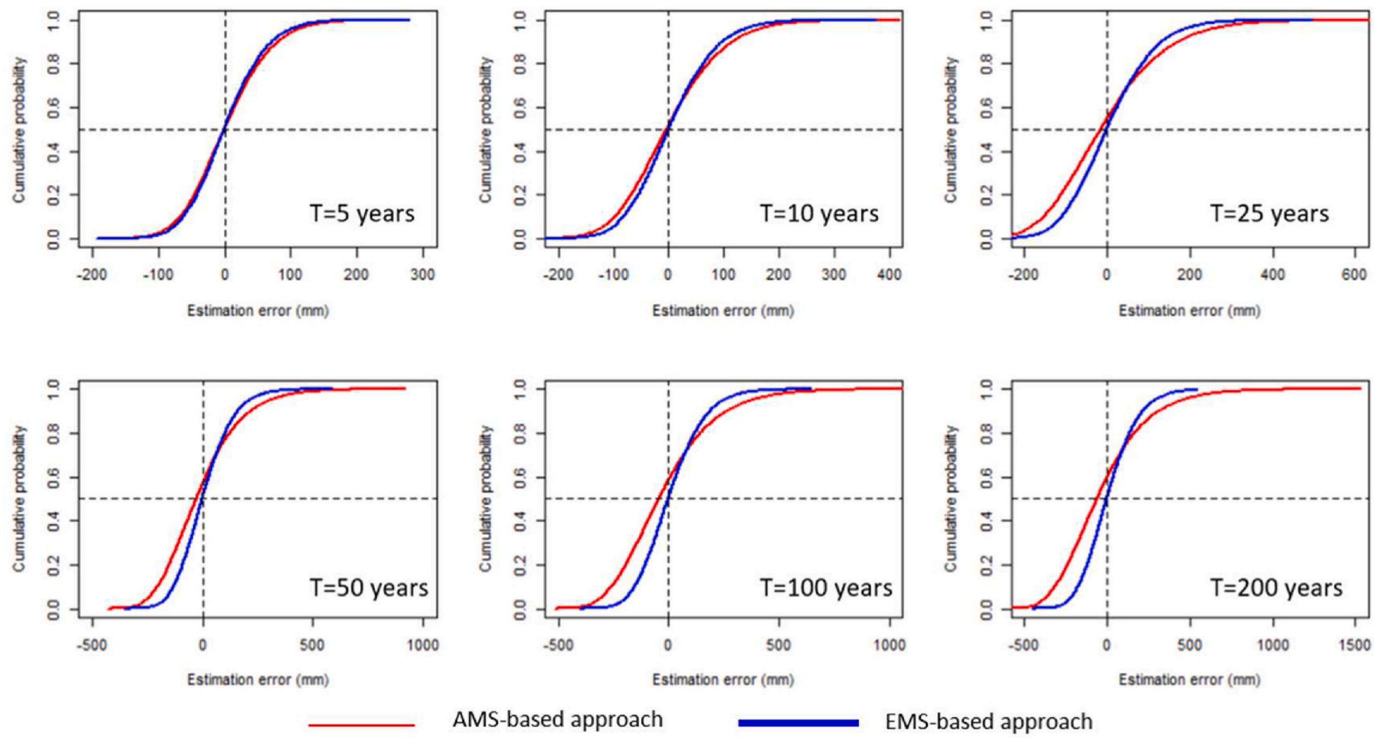

Fig. 7. Cumulative distribution functions of the T-year design rainfall estimation errors based on 10,000 block simulationruns.

3.3.3. *T-year AMR confidence intervals of the EMS-based and AMS-based approaches*

We divide the 10,000 blocks of T-year design rainfall estimates into 500 block groups, with each block group comprising 20 estimates of the T-year design rainfall. Each block group can then be considered as a random sample of size 20 of the T-year design rainfall. The T-year design rainfall represents the $\left(1-\frac{1}{T}\right)$ quantile of the AMR distribution. Sample quantiles of an absolutely continuous distribution are asymptotically normally distributed (Walker, 1968). Thus, the $(1-\alpha)100\%$ confidence interval (CI) of the theoretical T-year AMR (see Table 2) can be established by the following equation:

$$P\left(\overline{q}_T - t_{(1-\alpha/2),(n-1)}\frac{s_T}{\sqrt{n}} \le x_T \le \overline{q}_T + t_{(1-\alpha/2),(n-1)}\frac{s_T}{\sqrt{n}}\right) = 1-\alpha \quad (12)$$

where x_T represents the theoretical T-year AMR, n is the sample size, and $\overline{q}_T$ and s_T are the sample mean and standard deviation of the T-year AMS or EMS design rainfall estimates, respectively. Fig. 8 presents the T-year design rainfall CIs established using the AMS-based and EMS-based approaches. The theoretical T-year AMR covering rate of the 95% CIs was also calculated and is shown in Fig. 8. In general, the AMS-based approach yields wider 95% CIs than the EMS-based approach. As presented in Table 4, the average width of the 95% CIs of the EMS-based approach is 61%–90% of that of the AMS-based approach. In other words, for our simulation scenario, the EMS-based approach reduces 10%–39% of the uncertainty in design rainfall estimates as compared to the AMS-based approach. The theoretical T-year AMR covering rates for the EMS-based approach were approximately 0.942, whereas the covering rate of the 100, and 200-year AMR for the AMS-based approach were 0.906, and 0.894, respectively. These results indicate that the AMS-based approach yields design rainfall with higher uncertainty than the EMS-based approach.

4. EMS-based rainfall frequency analysis in Taiwan

The EMS-based approach has been applied to rainfall data of CWA-ARN stations in Taiwan. Hourly rainfall data over a 31-year (1987–2017) period are available at these stations. Quality control evaluation of the CWA-ARN rainfall data has been conducted, and the data are considered to be of high quality (Chen et al., 2016)). In this paper, we only present detailed results from four representative stations (north: Taipei, central: Taichung, south: Kaoshiung, and east: Hualien). The locations of these representative stations are shown in Fig. 2.

4.1. *EMS and AMS data extraction*

The proposed EMS-based approach requires the extraction of individual storm events from the historical hourly rainfall series. This is accomplished by considering the seasons of prevalent storm occurrences, event durations, average rainfall intensities, and the minimum inter-event time. Due to the commonly observed storm rainfall intermittence, a minimum inter-event time, which is the minimum time interval from the end of a storm event to the beginning of the next storm event, is required to separate individual storm events. The inter-event time varies among events, and its statistical characteristics also vary with prevalent storm types. However, the minimum inter-event time should be selected by considering the time of concentration of drainage basins. Rainfalls separated by a time span longer than the time of concentration of the drainage basin will not contribute jointly to runoff at the basin outlet. Given the short time of concentration for most rivers in Taiwan, a minimum inter-event time of 4 h was selected in this study.

Additionally, many stations identified storm events of 1-h durations with low rainfall amounts (<10 mm mostly) or those of longer durations with very low average rainfall intensities (<2 mm/h). These events were

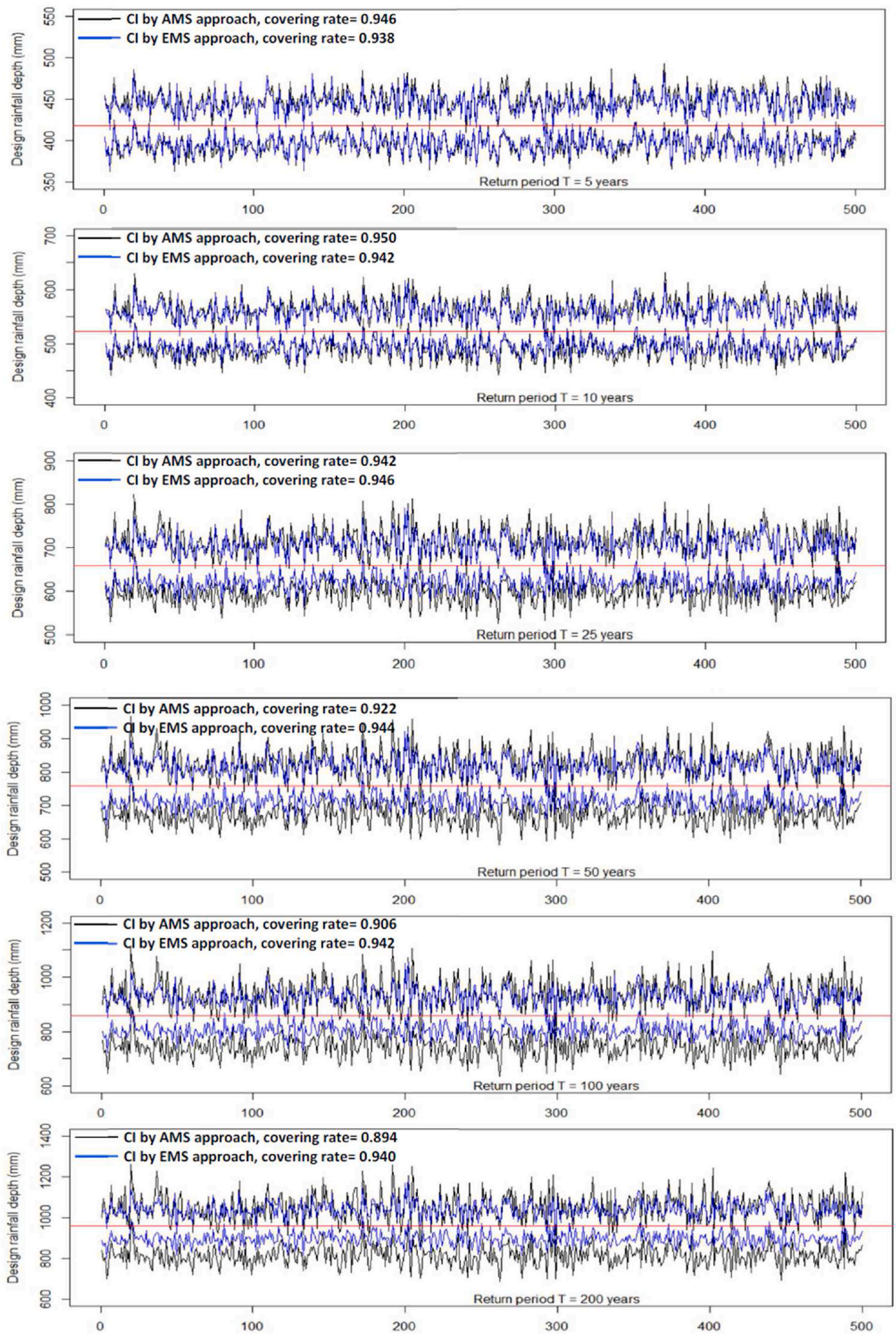

Fig. 8. *T*-year design rainfall 95% CIs established using the AMS-based and EMS-based approaches.

Table 4
Average width and covering rate of the 95% confidence intervals for the *T*-year design rainfalls.

	Return period *T* (years)					
	5	10	25	50	100	200
Average CI width (mm)						
AMS approach	53.90	78.13	120.08	156.06	194.29	234.11
EMS approach	48.66	65.49	88.68	106.72	125.03	142.97
EMS/AMS Ratio	0.90	0.84	0.74	0.68	0.64	0.61
T-year AMR covering rate						
AMS approach	0.946	0.950	0.942	0.922	0.906	0.894
EMS approach	0.938	0.942	0.946	0.944	0.942	0.940

excluded from subsequent analyses because they did not contribute to extreme rainfalls, and including these events in the EMS-based approach could significantly decrease the mean EMR values. Table 5 provides a summary of the criteria used for the extraction of different prevalent storm events. From each of the extracted storm events, the EMR of 1, 2, 3, 6, 12, 24, 48, and 72-h design durations were calculated. In addition, the traditional AMR of the *t*-hour design duration were extracted using a moving window of *t*-hour window size.

4.2. Characteristics of the prevalent storm events

The characteristics of the seasonal prevalent storms differ in terms of the occurrence frequency, storm duration, and total rainfall amounts. The characteristics of the prevalent storms recorded at the four representative stations are summarized in Table 6.

Meiyu storms and typhoons occur less frequently than summertime convective storms and WS frontal events; however, they have longer durations and higher event-total rainfalls. Compared with other stations, the Kaoshiung station recorded the least annual number of WS frontal events, and the Hualien station recorded the least annual number of Meiyu storm events. As a result, Kaoshiung receives very low rainfall from November to April, and Hualien receives the lowest Meiyu rainfall (see Fig. 2).

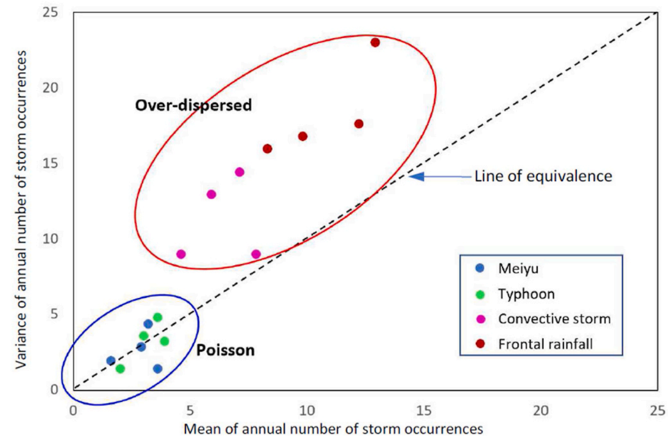

Fig. 9. Means and variances of the annual number of occurrences of different prevalent storms.

The annual number of prevalent storms plays a crucial role in the EMS-based frequency analysis. In Sections 3.2 and 3.3, the annual numbers of typhoons and Meiyu storms are assumed to follow the Poisson distribution. A unique property of the Poisson distribution is that its mean and variance are the same. The mean and variance of the annual number of prevalent storms at the four representative stations are plotted in Fig. 9. The annual number of typhoons and Meiyu storms can be well represented by the Poisson distribution. However, for summertime convective storms and WS frontal events, the variance of

Table 5
Criteria for separating storm events.

Prevalent storm type	Season of occurrence	Event duration d (hours)	Minimum inter-event-time (hours)	Minimum average rainfall intensity (mm/hr)
Meiyu	May to June (excluding the period of typhoon or tropical storm occurrences)	$d > 8$	4	2
Summertime convective storms	July to October (occurring after 12:00 p.m. and excluding the period of typhoon or tropical storm occurrences)	$d \leq 8$	4	2
Typhoons	Based on CWA record	$d > 8$	4	2
Winter-to-spring frontal events	November to April		4	2

Table 6
Summary of the prevalent storm characteristics at four representative stations.

Station		Annual count of events		Event duration (hours)		Event-total depth (mm)	
		Mean	Variance	Mean	Standard deviation	Mean	Standard deviation
Taipei (North)	Meiyu	2.9	2.9	19.6	13.3	80.9	66.0
	convective storms	12.9	23.0	3.6	2.0	28.0	25.2
	Typhoons	3.0	3.6	32.3	17.2	150.9	124.0
	WS frontal events	7.1	14.4	10.5	10.7	32.1	28.6
Taichung (Central)	Meiyu	3.6	1.4	22.1	15.3	89.9	72.4
	convective storms	9.8	16.8	3.8	2.1	23.8	22.0
	Typhoons	2.0	1.4	31.7	17.6	178.5	146.4
	WS frontal events	5.9	13.0	9.2	8.4	30.8	27.9
Kaoshiung (South)	Meiyu	3.2	4.41	19.9	14.8	106.2	113.5
	convective storms	12.2	17.6	2.9	2.1	17.4	17.7
	Typhoons	3.6	4.84	23.7	14.2	129.0	117.8
	WS frontal events	4.6	9.0	5.9	5.7	25.1	23.0
Hualien (East)	Meiyu	1.6	1.96	17.8	11.4	73.2	57.5
	convective storms	8.3	16.0	3.8	2.1	20.0	21.4
	Typhoons	3.9	3.24	30.9	16.6	189.6	151.7
	WS frontal events	7.8	9.0	7.0	8.3	26.5	54.3

Table 7
Fitted distribution parameters for the 1-h and 24-h EMRs (in mm) for different prevalent storm types recorded at four representative stations.

Taipei	1-h EMR PE3			24-h EMR PE3			Negative binomial		
	μ	σ	γ	μ	σ	γ	β	μ	σ^2
Meiyu	19.79	15.01	2.27	74.53	55.61	2.38	18.07	2.71	3.12
Convective	17.73	15.51	1.92	27.96	25.84	1.95	5.63	12.10	38.10
Typhoon	24.65	13.34	1.50	130.05	97.71	2.08	7.07	2.84	3.98
WS Frontal	9.91	6.78	1.99	30.21	24.06	1.48	5.35	6.77	15.34
Taichung									
Meiyu	22.59	13.55	1.69	81.76	55.68	1.85	12.40	3.48	4.46
Convective	15.61	13.54	2.02	23.82	21.92	2.07	18.64	9.77	14.90
Typhoon	29.76	21.51	2.13	155.18	124.61	1.98	100.00	1.90	1.94
WS Frontal	11.43	8.44	2.00	30.07	26.29	1.78	5.04	5.77	12.39
Kaoshiung									
Meiyu	25.29	17.86	1.97	92.55	83.31	2.51	10.21	3.23	4.24
Convective	11.87	10.60	2.10	17.36	17.55	2.26	31.72	12.19	16.88
Typhoon	27.81	18.56	1.61	118.90	100.70	2.22	13.10	3.61	4.61
WS Frontal	12.50	9.08	1.75	25.07	22.81	2.05	5.79	4.61	8.29
Hualien									
Meiyu	16.86	12.53	2.03	68.49	53.54	2.47	8.56	1.55	1.83
Convective	12.18	11.90	2.61	19.98	20.79	2.55	14.70	8.32	13.03
Typhoon	30.47	19.38	1.29	155.90	111.04	1.29	100.00	3.74	3.88
WS Frontal	8.83	7.24	2.33	25.40	37.85	3.71	73.16	7.90	8.76

Note: Parameters μ,σ,and γ are the mean, standard deviation, and coefficient of skewness, respectively. Parameter β denotes the size of the negative binomial distribution.

the annual number of storms is apparently higher than would be expected by the Poisson distribution for all four stations. This is known as the *over-dispersion* of count data.

Various methods of modeling the overdispersed count data have been reported in the literature (Johnson et al., 2005; Lloyd-Smith, 2007; Lindén and Mäntyniemi, 2011; Xekalaki, 2014). In this study, we used the Poisson-gamma mixture approach to derive the negative binomial distribution for the overdispersed data. If a random variable X has a Poisson distribution with parameter λ which itself has a gamma distribution with scale parameter α and shape parameter β, then the distribution of X will be a negative binomial distribution with mean $\alpha\beta$ and dispersion parameter β. The derivation of the negative binomial distribution modeling of the overdispersed data is detailed in Appendix B. Because the Poisson distribution can be considered a special case of the negative binomial distribution, the annual numbers of all prevalent storm types were modeled using the negative binomial distribution in our study.

The EMR and traditional AMR for the different storm types were fitted to the Pearson type III distribution. The R package fitdistrplus (Delignette-Muller and Dutang, 2015) was used for the maximum-likelihood parameter estimation of the negative binomial and Pearson type III distributions. Table 7 presents the fitted parameters of 1-h and 24-h EMR for different prevalent storms at the four representative stations. In general, the means and variances of the annual number of storm events (values in the last two columns of Table 7) were similar for typhoons and Meiyu, indicating that the application of negative binomial distribution modeling was suitable for both the Poisson-distributed and overdispersed count data.

4.3. AMR contributing probabilities of individual storm types – representative stations

By using the parameters listed in Table 7, the CDFs of the overall and storm-type-specific AMR $(G, F^*_{Mei}, F^*_{Con}, F^*_{Typ}, F^*_{Frn})$ were derived and are presented in Fig. 10. Similar to the AMR contributing probabilities described in Section 3.3.1, the overall AMR contributing probabilities of different storm types at the four representative stations were calculated using the inverse probability integral transformation method. Table 8 presents the AMR contributing probabilities for the short (1-h) design duration and long (24-h) design duration. A comparison of the CDFs in Fig. 10 and AMR contributing probabilities in Table 8 reveals the following insightful observations:

(1) The 24-h AMR is mostly caused by typhoons. This is particularly true for Hualien station, which has an AMR contributing probability of 0.83. This is because nearly 60% of typhoon landfalls occurred near the Hualien station (CWA, 2022).
(2) The 24-h AMR contributing probabilities of Meiyu storms are higher in the southern (Kaoshiung, 0.35) and central (Taichung, 0.34) regions than in the northern (Taipei, 0.23) and eastern (Hualien, 0.06) regions. Meiyu fronts move from southern China to Taiwan, and Hualien is located on the lee side of the Central Mountain Range that stretches from north to south, with an elevation ranging from 2,000 to 3,900 m above the mean sea level. Thus, among the four representative stations, Hualien station receives the least amount of Meiyu rainfall (see Fig. 2) and has the lowest AMR contributing probability of Meiyu storms.
(3) Regarding the 24-h AMR exceeding the 5-year return level, typhoons are the near-dominant storm type for Taipei, Taichung, and Hualien stations. However, the contribution of Meiyu storms is nonnegligible at Kaoshiung station.
(4) The major storm types that contribute to the 1-h AMR vary among stations. Uniquely, for the 1-h AMR exceeding the 5-year return level, convective storms have a dominant contribution for Taipei station, with 1-h AMR contributing probability of 0.60. By contrast, typhoons are the major contributors of the 1-h AMR for other stations. Taipei, a highly populated city situated on a basin plain in northern Taiwan, is nearly surrounded by mountains, and whether the urban heat island effect has contributed to the high dominance of convective storms in Taipei requires further study.

Understanding the AMR contributing probabilities of different storm types is important for applications such as reservoir management and flood risk mitigation. In West Coast in the United States, long-duration storms tend to generate the largest amounts of precipitation, potentially contributing more to water resources or leading to extreme floods (Lamjiri et al., 2017). The results presented in this section indicate that the EMS-based approach can offer a comprehensive understanding of AMR contributing probabilities of different storm types which may be helpful for reservoir management and flood risk mitigation.

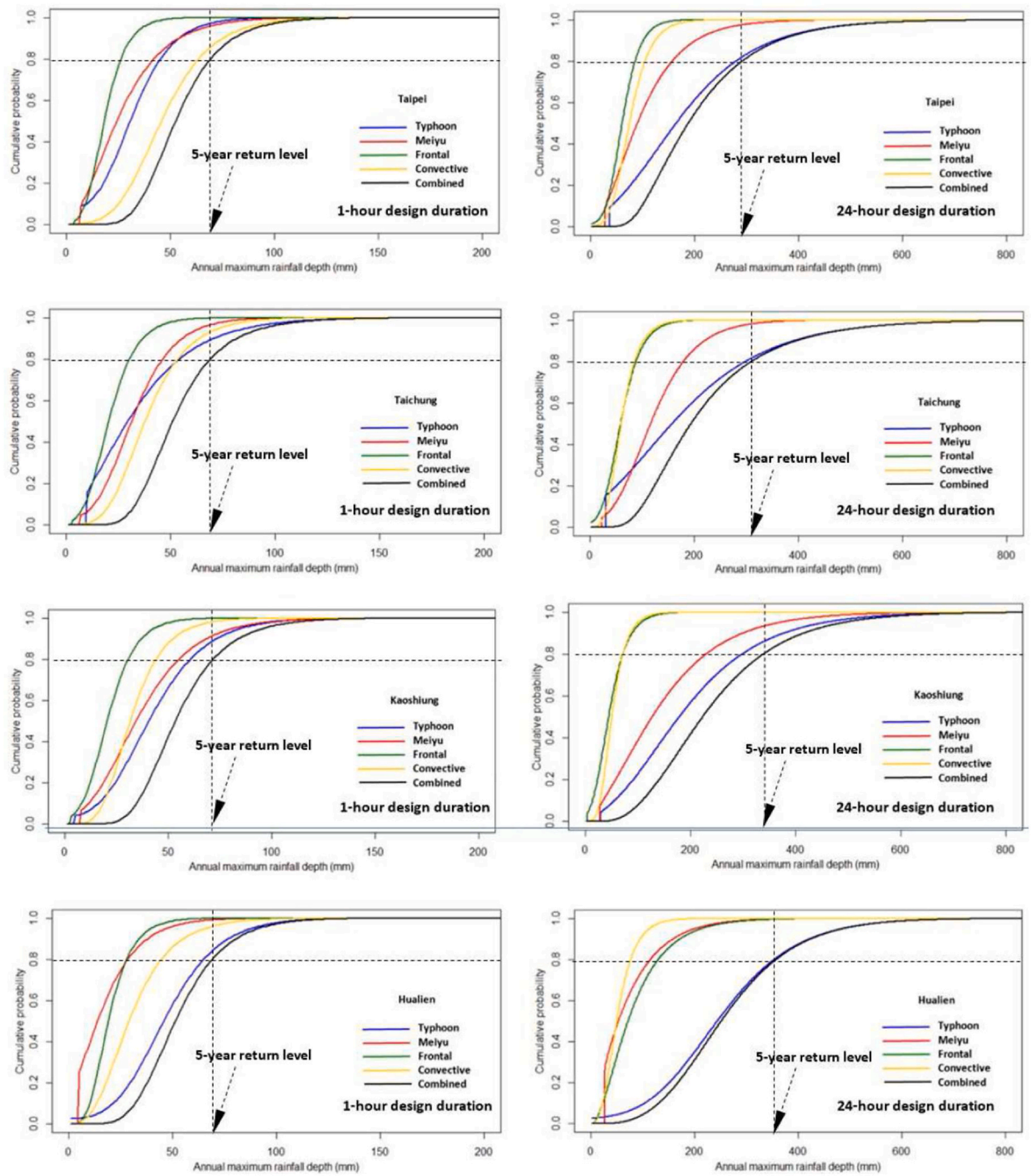

Fig. 10. Cumulative distribution functions of the overall and storm-type-specific annual maximum rainfalls at four representative stations.

Table 8
Overall AMR contributing probabilities of different prevalent storm types recorded at four representative stations.

	Meiyu	Convective	Typhoon	WS Frontal
Taipei				
1-h AMR	0.17	0.60	0.21	0.03
24-h AMR	0.23	0.09	0.64	0.04
Taichung				
1-h AMR	0.25	0.40	0.29	0.06
24-h AMR	0.34	0.05	0.56	0.05
Kaoshiung				
1-h AMR	0.32	0.19	0.44	0.05
24-h AMR	0.35	0.02	0.61	0.02
Hualien				
1-h AMR	0.06	0.23	0.66	0.05
24-h AMR	0.06	0.02	0.83	0.08

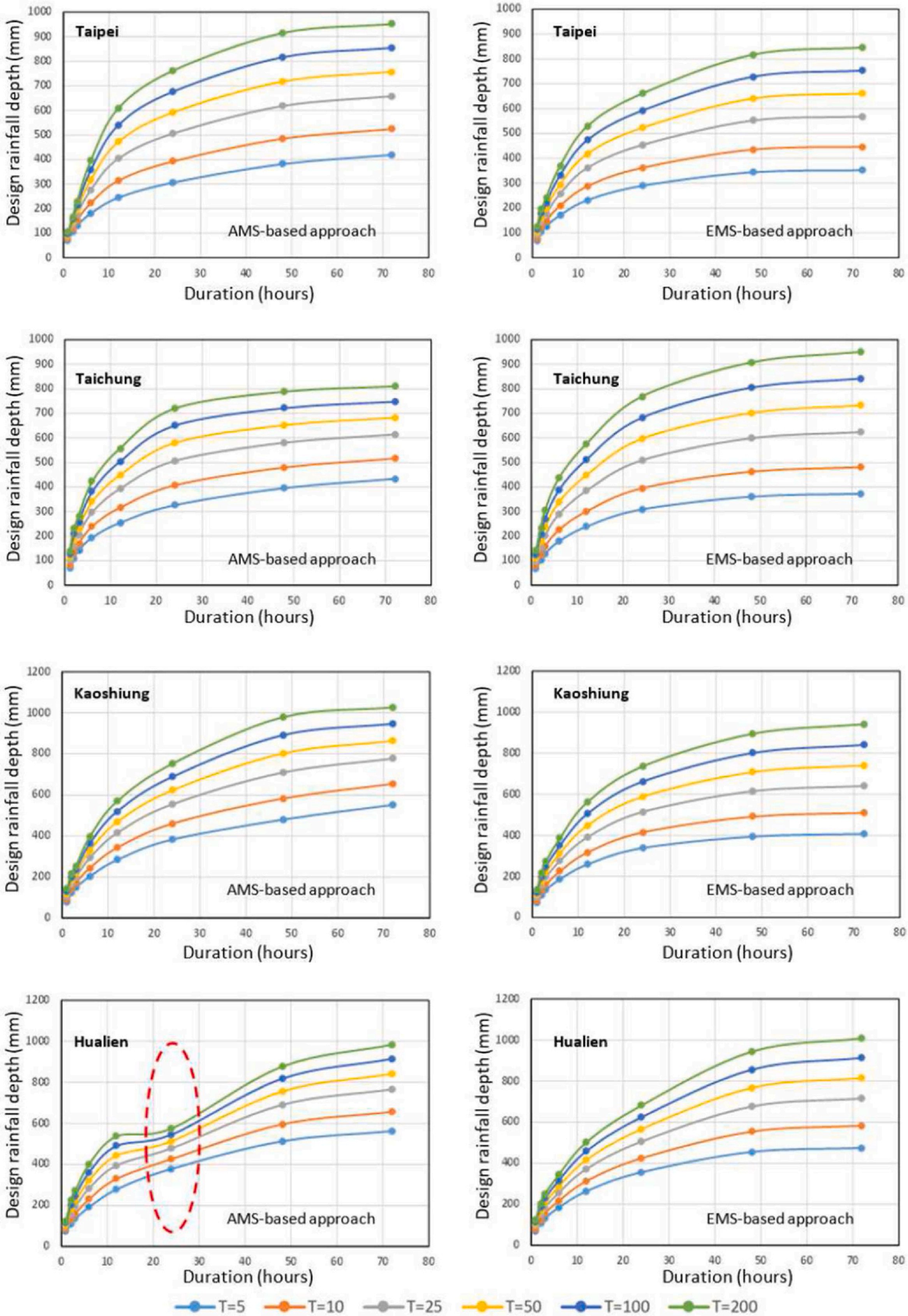

Fig. 11. Design rainfall depths for the AMS-based and EMS-based approaches.

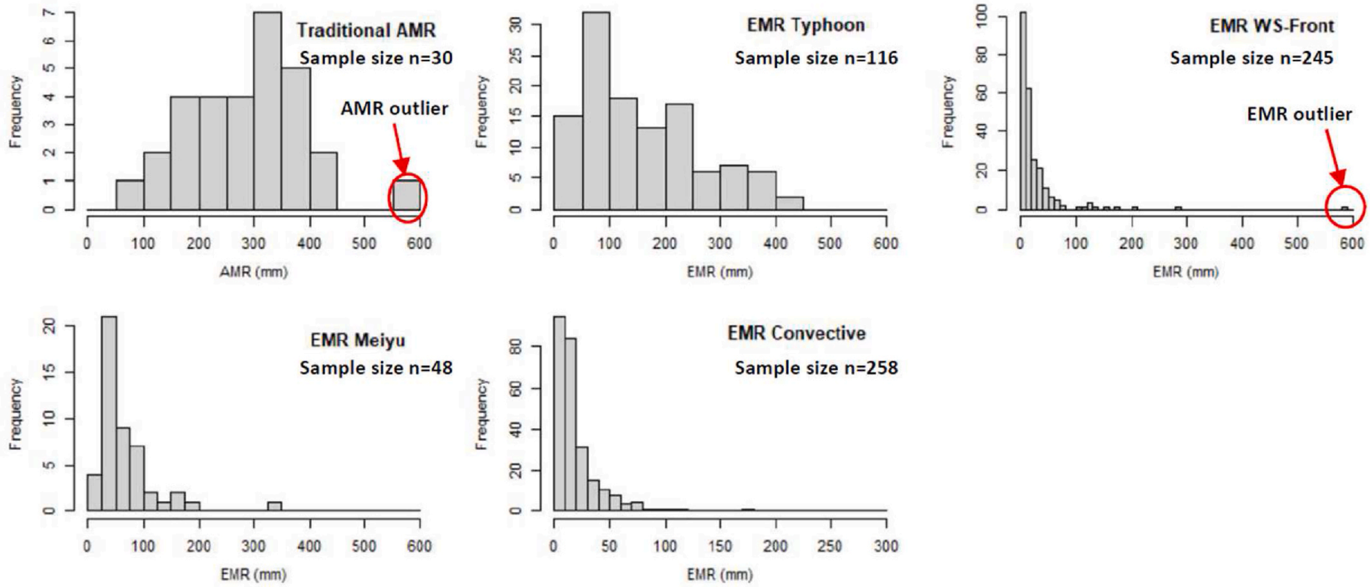

Fig. 12. Histograms of the 24-h traditional AMR and EMR of different prevalent storm types at the Hualien station. The red circle shows an AMR and EMR outlier caused by a WS frontal event on Nov. 9, 1996. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

4.4. Comparison of the EMS-based and AMS-based design rainfalls

By using the parameters listed in Table 7 and the mixture distribution modeling approach described in Section 3.2, the EMS-based design rainfalls were calculated. The AMS-based design rainfalls were also calculated based on the fitted PE3 distribution of the traditional AMRs. Fig. 11 demonstrates the depth-duration-frequency (DDF) curves of the AMS-based and EMS-based approaches at the four representative stations. An unusual pattern (see the dashed ellipse in Fig. 11)was observed in the DDF curves of Hualien station. Compared with the design rainfalls of other design durations, the 24-h design rainfalls of the AMS-based approach seemed to vary in a relatively narrow range.

To investigate the reason underlying the unusual pattern, we examined the histograms of the 24-h traditional AMR and the 24-h EMR for different storm types at Hualien station. As presented in Fig. 12, the traditional AMR includes an outlier (586 mm) caused by a WS frontal event on November 9, 1996. The 24-h traditional AMR of Hualien station has a standard deviation of 102.67 mm, which is lower than the corresponding standard deviations of Taipei (119.51 mm), Taichung (124.61 mm), and Kaoshiung (125.08 mm) stations. If the AMR outlier is removed, the standard deviation of the 24-h traditional AMR of Hualien station becomes 88.06 mm, and the AMR data becomes negatively skewed (coefficient of skewness $= -0.264$). The smaller standard deviation of the 24-h traditional AMR caused the 24-h design rainfall of Hualien station to vary in a narrow range, and the AMR of higher return periods ($T \geq 25$) could have been underestimated. By contrast, the EMS-based DDF curves in Fig. 11 do not show similar unusual behavior. The sample size of the rainfall data series plays a crucial role in frequency analysis. In our study, the AMS-based approach used only 30 traditional AMRs, whereas the EMS-based approach used 667 EMRs. A larger sample size can reduce the parameter uncertainty; thus, the EMS-based approach is more robust than the AMS-based approach.

4.5. Potential for rainfall frequency analysis under climate change

The EMS-based approach considers the statistical characteristics of the annual number of storms and the EMR for different prevalent storm types. Several studies have found that changes in heavy precipitation under climate change are linked to changes in the occurrence frequency, intensity, and total rainfall of storm events (Myhre et al., 2019; Bevacqua et al., 2020; Martel et al., 2021; Gründemann et al., 2022). Therefore, understanding these characteristics of different storm types is crucial for conducting rainfall frequency analysis under climate change. The traditional AMS-based frequency analysis does not consider these storm characteristics and therefore cannot offer detailed insights into the occurrence frequencies and AMR contributing probabilities of individual prevalent storm types.

New generation of convection-permitting climate models (CPMs) better represent land-surface characteristics and small-scale processes in the atmosphere, such as convection (Kendon et al., 2021). They are able to better simulate rainfall extremes of short time scales (up to sub-hourly scales) with kilometer-scale spatial resolution (Vergara-Temprado et al., 2021; Martel et al., 2021). CPM future projections of short time-scale rainfall extremes are crucial to generating IDF curves and extreme value modeling of future rainfall (Kourtis and Tsihrintzis, 2022). However, large CPM ensembles that comprehensively sample future uncertainties are costly. Kourtis and Tsihrintzis (2022) indicated that the only way to estimate future IDF curves is using future climate projections from a single model with a rather short reference period, for example, 10–20 years. However, using a single model output with short reference period will inevitably lead to high uncertainty in the AMS-based rainfall frequency analysis. This difficulty can be overcome by using the proposed EMS-based approach, which considers the storm characteristics of different storm types and involves a large sample size for EMR modeling. Therefore, the EMS-based approach is well-suited for rainfall frequency analysis under climate change using output from a single CPM.

Recent studies indicate that design rainfalls of shorter durations and longer return periods are likely to have a higher degree increase in a warmer climate (Martel et al., 2021). Such changes will significantly impact stormwater management in cities and in small rural catchments. Most existing stormwater drainage infrastructures, which were designed using IDF reationships established based on decades-old historical data, are ill-adapted to climate change. Thus, climate change adaptation requires IDF updating to account for an evolving climate and reduce infrastructure vulnerability. Although building climate-resilient stormwater management infrastructure requires a larger investment in the initial construction cost, it will ultimately lead to a lower total cost over the long lifespan of the infrastructure elements (Martel et al., 2021).

5. Conclusions

This study proposed an event-based mixture distribution modeling approach for rainfall frequency analysis by using the EMR. The novelty of using EMR of individual storm events of all storm types largely increases the sample size for parameter estimation and reduces the uncertainty of the design rainfall estimates. It also enables a thorough investigation of the AMR contributing probabilities of different storm types. The EMS-based approach is particularly suitable for regions with cyclostationary climate patterns and rainfall stations with short record lengths. A rigorous stochastic simulation demonstrated that the EMS-based approach was superior to the traditional AMS-based approach. The EMS-based approach considers the characteristics of all prevalent storm types and can be used to elucidate the contributions of individual storm types to the AMR. Concluding remarks are as follows:

1. Unlike the AMS and POT series, the EMS is considered to be a complete series because it consists of the EMRs of all individual storm events, including all prevalent storm types.
2. The design rainfall estimates of the EMS-based approach are nearly unbiased and have smaller RMSE than the design rainfall estimates of the AMS-based approach. The average widths of the 95% CI of the *T*-year AMR of the EMS-based approach are also smaller than those of the AMS-based approach.
3. Among the four prevalent storm types in Taiwan, the annual numbers of Meiyu storms and typhoons can be well described by the Poisson distribution, whereas the annual numbers of the summertime convective storms and the WS frontal events are overdispersed. The negative binomial distribution can be used to model both the Poisson-distributed and overdispersed count data.
4. The 24-h AMR is mostly caused by typhoons at all representative stations. However, the contribution of Meiyu storms to the AMR of high quantiles is nonnegligible at the Kaoshiung station.
5. The 24-h AMR contributing probabilities of Meiyu storms are higher in the southern and central regions than in the northern and eastern regions of Taiwan. Particularly, the AMR contributing probability of Meiyu storms is the lowest at the Hualien station due to its geographic location.
6. Convective storms have a dominant contribution to the 1-h AMR at the Taipei station. By contrast, typhoons are the major contributor to the 1-h AMR at the other stations.
7. The sample size of the rainfall data series is a crucial element for frequency analysis. A larger sample size can reduce the parameter uncertainty; thus, the EMS-based approach is more robust than the AMS-based approach.
8. The proposed EMS-based approach considers the characteristics of different storm types and involves a large sample size for EMR modeling; therefore, it is well-suited for rainfall frequency analysis under climate change using output from a single CPM.

CRediT authorship contribution statement

Ke-Sheng Cheng: Conceptualization, Formal analysis, Funding acquisition, Methodology, Supervision, Writing – original draft, Writing – review & editing. **Bo-Yu Chen:** Data curation, Formal analysis, Software, Validation. **Teng-Wei Lin:** Data curation, Formal analysis, Software, Validation. **Kimihito Nakamura:** Conceptualization, Funding acquisition, Resources. **Piyatida Ruangrassamee:** Conceptualization, Funding acquisition, Resources. **Hidetaka Chikamori:** Conceptualization, Methodology, Resources.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Kesheng Cheng reports financial support was provided by National Science and Technology Council, Taiwan.

Data availability

Data will be made available on request.

Acknowledgments

We acknowledge the funding support of the National Science and Technology Council, Taiwan (MOST 107-2119-M-002-048 -) and the Office of International Affairs, National Taiwan University (2020-2021 International Collaboration Seed Funding Program). The first author is grateful to Kyoto University, Japan, and Chulalongkorn University, Thailand, for hosting his academic visits.

Appendix A. Gamma distribution parameter estimation by using the method of *L*-moments (Hosking and Wallis, 1997; Rao and Hamed, 2000; Liou et al., 2008)

A gamma random variable has the following probability density function

$$f(x;\alpha,\beta)=\frac{1}{\alpha\Gamma(\beta)}\left(\frac{x}{\alpha}\right)^{\beta-1}e^{-x/\alpha},\alpha>0,\beta>0,x\geq 0. \tag{A.1}$$

Given a random sample $\{x_1,x_2,\cdots,x_n\}$, the sample L-moments ($\ell_r, r=1,2,3,4$) and sample L-moment ratios (t_r) can be calculated by

$$\ell_1=b_0 \tag{A.2}$$

$$\ell_2=2b_1-b_0 \tag{A.3}$$

$$\ell_3=6b_2-6b_1+b_0 \tag{A.4}$$

$$\ell_4=20b_3-30b_2+12b_1-b_0 \tag{A.5}$$

where b_r represents the unbiased estimator of the probability weighted moment and is given by

$$b_r=\frac{1}{n}\sum_{j=r+1}^{n}\frac{(j-1)(j-2)\cdots(j-r)}{(n-1)(n-2)\cdots(n-r)}x_{j:n} \tag{A.6}$$

In the above equation $x_{j:n}$ represents the k-th order statistic from a random sample of size n. The method of L-moment estimator of the scale parameter

α and shape parameter β can be calculated as follows, where $t = \ell_2/\ell_1$.

If $0 < t < \frac{1}{2}$,

$$z = \pi t^2 \tag{A.7}$$

$$\hat{\beta} = \frac{(1 - 0.308z)}{(z - 0.05812z^2 + 0.01765z^3)} \tag{A.8}$$

If $\frac{1}{2} \leq t < 1$,

$$z = 1 - t \tag{A.9}$$

$$\hat{\beta} = \frac{(0.7213z - 0.5947z^2)}{(1 - 2.1817z + 1.2113z^2)} \tag{A.10}$$

$$\hat{\alpha} = \ell_1/\hat{\beta} \tag{A.11}$$

Several R packages are available for sample *L*-moments calculation, including lmom, lmomco, and Lmoments.

Appendix B. Derivation of the negative binomial distribution for the Poisson-gamma mixture process

The negative binomial distribution can be used to model the number of failures in a sequence of independent and identically distributed Bernoulli trials before the occurrence of the *r*th success. Let *p* be the probability of success for each Bernoulli trial, then, the number of failures before achieving ***r*** successes, denoted by *N*, has the following probability mass function,

$$f_N(n; r, p) \equiv \Pr(N = n) = \binom{n + r - 1}{n} (1 - p)^n p^r$$

$$= \frac{\Gamma(n + r)}{\Gamma(n+1)\Gamma(r)} (1 - p)^n p^r, n = 0, 1, 2, \ldots \tag{B.1}$$

$$E(N) = \frac{r(1 - p)}{p}, Var(N) = \frac{r(1 - p)}{p^2} \tag{B.2}$$

where ***r*** is called the size parameter of the negative binomial distribution. Let *X* be a Poisson random variable with probability mass function

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, \ldots \tag{B.3}$$

Suppose that the Poisson parameter λ varies and has the following gamma distribution with scale parameter α and shape parameter β,

$$f(\lambda) = \frac{1}{\alpha\Gamma(\beta)} \left(\frac{\lambda}{\alpha}\right)^{\beta - 1} e^{-\left(\frac{\lambda}{\alpha}\right)} \tag{B.4}$$

For the above Poisson-gamma mixture process, the probability mass function of *X* is expressed by

$$p(x) = \int_0^{+\infty} p(x|\lambda) f(\lambda) d\lambda = \int_0^{+\infty} \frac{e^{-\lambda}\lambda^x}{x!} \frac{1}{\alpha\Gamma(\beta)} \left(\frac{\lambda}{\alpha}\right)^{\beta - 1} e^{-\left(\frac{\lambda}{\alpha}\right)} d\lambda$$

$$= \frac{1}{\alpha^\beta x! \Gamma(\beta)} \int_0^{+\infty} \lambda^{(x+\beta-1)} e^{-\lambda\left(1 + \frac{1}{\alpha}\right)} d\lambda \tag{B.5}$$

From the gamma density, a gamma distribution with scale parameter $\left(1 + \frac{1}{\alpha}\right)$ and shape parameter $(\boldsymbol{x} + \beta - 1)$ yields

$$\int_0^{+\infty} \frac{1}{\left(\frac{\alpha}{1+\alpha}\right)\Gamma(x + \beta)} \left(\frac{\lambda}{[\alpha/(1 + \alpha)]}\right)^{x+\beta-1} e^{-\left(\frac{\lambda}{[\alpha/(1+\alpha)]}\right)} d\lambda$$

$$= \left(\frac{1 + \alpha}{\alpha}\right)^{x+\beta} \frac{1}{\Gamma(x + \beta)} \int_0^{+\infty} \lambda^{(x+\beta-1)} e^{-\left(\frac{\lambda}{[\alpha/(1+\alpha)]}\right)} d\lambda = 1 \tag{B.6}$$

Therefore,

$$\int_0^{+\infty} \lambda^{(x+\beta-1)} e^{-\lambda\left(1 + \frac{1}{\alpha}\right)} d\lambda = \left(\frac{\alpha}{1 + \alpha}\right)^{x+\beta} \Gamma(x + \beta) \tag{B.7}$$

$$p(x) = \frac{1}{\alpha^{\beta} x! \Gamma(\beta)} \left(\frac{\alpha}{1+\alpha}\right)^{x+\beta} \Gamma(x+\beta) = \frac{\Gamma(x+\beta)}{x!\Gamma(\beta)} \left(\frac{\alpha}{1+\alpha}\right)^{x} \left(\frac{1}{1+\alpha}\right)^{\beta} \tag{B.8}$$

The above equation represents the probability mass function of a negative binomial distribution with the probability of success $\left(\frac{1}{1+\alpha}\right)$ and size (or dispersion) parameter β. The mean and variance of the above negative binomial distribution are

$$E(X) = \frac{\beta\left(\frac{\alpha}{1+\alpha}\right)}{\left(\frac{1}{1+\alpha}\right)} = \alpha\beta, Var(X) = \frac{\beta\left(\frac{\alpha}{1+\alpha}\right)}{\left(\frac{1}{1+\alpha}\right)^2} = (1+\alpha)\alpha\beta. \tag{B.9}$$

The variance of the negative binomial distribution can also be expressed by

$$Var(X) = E(X) + \frac{[E(X)]^2}{\beta} \tag{B.10}$$

$$\text{If } \frac{[E(X)]^2}{\beta} \ll E(X), Var(X) \approx E(X) \tag{B.11}$$

Thus, as the size parameter β increases, the variance converges to the expected value, and the negative binomial distribution turns into a Poisson distribution.

References

Akaike, H., 1974. A new look at statistical model identification. IEEE Trans Autom Control AC- 19, 716–723.

Allamanno, P., Laio, F., Claps, P., 2011. Effects of disregarding seasonality on the distribution of hydrological extremes. Hydrol. Earth Syst. Sci. 15, 3207–3215.

Ashkar, F., Rousselle, J., 1987. Partial duration series modelling under the assumption of a Poissonian flood count. J. Hydrol. 90 (1–2), 135–144.

Bačová-Mitková, V., Onderka, M., 2010. Analysis of extreme hydrological events on the Danube using the peak over threshold method. J. Hydrol. Hydromechanics 58 (2), 88–101. https://doi.org/10.2478/v10098-010-0009-x.

Bevacqua, E., Vousdoukas, M.I., Zappa, G., et al., 2020. More meteorological events that drive compound coastal flooding are projected under climate change. Commun. Earth Environ. 1, 47. https://doi.org/10.1038/s43247-020-00044-z.

Bezak, V., Brilly, M., Šraj, M., 2014. Comparison between the peaks-over-threshold method and the annual maximum method for flood frequency analysis. Hydrol. Sci. J. 59 (5), 959–977. https://doi.org/10.1080/02626667.2013.831174.

Chang, C.W., Chiang, C.T., Liu, K.Y., Su, S.H., 2019. The comparison of objective diagnosis methods for Taiwan frontal system classification. Atmos. Sci. 47 (1), 1–20 (in Chinese).

Chen, G.T.J., Chou, H.C., 2006. A Summertime severe weather event occurring in the Taipei basin. Terr. Atmos. Ocean Sci. 17 (1), 3–22.

Chen, P.Y., Shen, L.Y., Chen, Y.L., 2016. Analysis of rainfall exceedance probability for CWA-ARN stations (in Chinese). In: 2016 Weather Analysis and Forecast Conference. Taipei, Taiwan.

Cheng, K.S., Lin, Y.C., Liou, J.J., 2008. Rain-gauge network evaluation and augmentation using geostatistics. Hydrol. Process. 22, 2554–2564.

Cheng, K.S., Hou, J.C., Liou, J.J., Wu, Y.C., Chiang, J.L., 2011. Stochastic simulation of bivariate gamma distribution: a frequency-factor based approach. Stoch. Environ. Res. Risk Assess. 25, 107–122. https://doi.org/10.1007/s00477-010-0427-7.

Claps, P., Liao, F., 2003. Peak over threshold analysis of flood and rainfall frequency curves. In: Hydrological Risk: Recent Advances in Peak River Flow Modelling, Prediction and Real-Time Forecasting. Assessment of the Impacts of Land-Use and Climate Changes. Proceedings of the ESF LESC Exploratory Workshop, Bologna, Italy, pp. 24–25. October 2003.

Coles, S., 2001. An Introduction to Statistical Modelling of Extreme Values. Springer, N. Y.

Coles, S., Davison, A., 2008. Statistical Modelling of Extreme Values. http://stat.epfl.ch.

CWA, 2022. 100 Questions about Typhoons. Central Weather Agency, Taiwan, ROC (in Chinese).

CWA, 2023. Region-wise Seasonal Weather Patterns in Taiwan. https://www.cwa.gov.tw/Data/service/hottopic/14277057430.pdf. (Accessed 30 November 2023) (in Chinese). Central Weather Agency, Taiwan, ROC.

D'Agostino, R.B., Stephens, M.A., 1986. Goodness-of-fit Techniques. Marcel Dekker, New York.

Deidda, R., Hellies, M., Langousis, A., 2021. A critical analysis of the shortcomings in spatial frequency analysis of rainfall extremes based on homogeneous regions and a comparison with a hierarchical boundaryless approach. Stoch. Environ. Res. Risk Assess. 35, 2605–2628. https://doi.org/10.1007/s00477-021-02008-x.

Delignette-Muller, M., Dutang, C., 2015. fitdistrplus: an R package for fitting distributions. J. Stat. Software 64 (4), 1–34.

El-Aroui, M.-A., Diebolt, J., 2002. On the use of the peaks over thresholds method for estimating out-of-sample quantiles. Comput. Stat. Data Anal. 39, 453–475.

Gründemann, G.J., van de Giesen, N., Brunner, L., et al., 2022. Rarest rainfall events will see the greatest relative increase in magnitude under future climate change. Commun. Earth Environ. 3, 235. https://doi.org/10.1038/s43247-022-00558-8.

Gumbel, E.J., 1954. In: Statistical Theory of Extreme Values and Some Practical Applications. Applied Mathematics Series, 33 (first ed. U.S. Department of Commerce, National Bureau of Standards.

Gumbel, E.J., 1963. Statistics of Extremes. Columbia University Press, New York, NY.

Hosking, J.R.M., Wallis, J.R., 1997. Regional Frequency Analysis: an Approach Based on L-Moments. Cambridge University Press, Cambridge.

Huxham, S.H., Mcgilchrist, C.A., 1969. On the extreme value distribution for describing annual flood series. Water Resour. Res. 5, 1404–1405.

Johnson, N.L., Kemp, A.W., Kotz, S., 2005. Univariate Discrete Distributions, third ed. Wiley & Sons, Inc.

Karim, F., Hasan, M., Marvanek, S., 2017. Evaluating annual maximum and partial duration series for estimating frequency of small magnitude floods. Water 9, 481. https://doi.org/10.3390/w907048.

Kendon, E.J., Prein, A.F., Senior, C.A., Stirling, A., 2021. Challenges and Outlook for Convection-Permitting Climate Modelling, vol. 379. Philosophical Transaction Royal Society A, 2019054720190547. https://doi.org/10.1098/rsta.2019.0547.

Kim, K.Y., 2000. Statistical prediction of cyclostationary processes. J. Clim. 13, 1098–1115.

Kourtis, I.M., Tsihrintzis, V.A., 2022. Update of intensity-duration-frequency (IDF) curves under climate change: a review. Water Supply 22 (5), 4951–4974. https://doi.org/10.2166/ws.2022.152.

Lamjiri, M.A., Dettinger, M.D., Ralph, F.M., Guan, B., 2017. Hourly storm characteristics along the U.S. West Coast: role of atmospheric rivers in extreme precipitation. Geophys. Res. Lett. 44, 7020–7028. https://doi.org/10.1002/2017GL074193.

Langbein, W.B., 1949. Annual floods and the partial-duration flood series. Trans. Am. Geophys. Union 30 (6), 879–881. https://doi.org/10.1029/TR030i006p00879.

Lee, O., Sim, I., Kim, S., 2020. Application of the non-stationary peak-over-threshold methods for deriving rainfall extremes from temperature projections. J. Hydrol. 585, 124318 https://doi.org/10.1016/j.jhydrol.2019.124318.

Lindén, A., Mäntyniemi, M., 2011. Using the negative binomial distribution to model overdispersion in ecological count data. Ecology 92 (7), 1414–1421.

Liou, J.J., Wu, Y.C., Cheng, K.S., 2008. Establishing acceptance regions for L-moments-based goodness-of-fit test by stochastic simulation. J. Hydrol. 355, 49–62.

Lloyd-Smith, J.O., 2007. Maximum likelihood estimation of the negative binomial dispersion parameter for highly overdispersed data, with applications to infectious diseases. PLoS One 2 (2), e180. https://doi.org/10.1371/journal.pone.0000180. PMID: 17299582; PMCID: PMC1791715.

Madsen, H., Rosbjerg, D., Harremoes, P., 1993. Application of the partial duration series approach in the analysis of extreme rainfalls. In: Kundzewicz, Z.W., et al. (Eds.), Extreme Hydrological Events: Precipitation, Floods and Droughts, vol. 213. IAHS Press, IAHS Publ, Wallingford, pp. 257–366.

Madsen, H., Rosbjerg, D., 1997. The partial duration series method in regional index-flood modelling. Water Resour. Res. 33 (4), 737–746.

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

Martel, J.-L., Brissette, F.P., Lucas-Picher, P., et al., 2021. Climate change and rainfall intensity–duration–frequency curves: overview of science and guidelines for adaptation. J. Hydrol. Eng., ASCE 26 (10), 03121001. https://doi.org/10.1061/(ASCE)HE.1943-5584.0002122.

Mascaro, G., 2018. On the distributions of annual and seasonal daily rainfall extremes in central Arizona and their spatial variability. J. Hydrol. 559, 266–281.

Myhre, G., Alterskjær, K., Stjern, C.W., et al., 2019. Frequency of extreme precipitation increases extensively with event rareness under global warming. Sci. Rep. 9, 16063 https://doi.org/10.1038/s41598-019-52277-4.

Naghettini, M. (Ed.), 2017. Fundamentals of Statistical Hydrology. Springer.

Rao, A.R., Hamed, K.H., 2000. Flood Frequency Analysis. CRC Press.

Rogencamp, G., Barton, J., 2012. The Lockyer Creek Flood of January 2011: what Happened and How Should We Manage Hazard for Rare Floods. 52nd Annual

Floodplain Management Association Conference. http://www.floodplainconference.com/papers2012.php.
Rustomji, P., Bennett, N., Chiew, F., 2009. Flood variability east of Australia's great dividing range. J. Hydrol. 374, 196–208.
Sargood, M.B., Cohen, T.J., Thompson, C.J., Croke, J., 2015. Hitting rock bottom: morphological responses of bedrock-confined streams to a catastrophic flood. Earth Surf. Dyn. 3 (2), 265–279.
Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat. 6 (2), 461–464.
Stedinger, J.R., Vogel, R.M., Foufoula-Georgiou, E., 1993. Frequency analysis of extreme events. In: Maidment, D.R. (Ed.), Handbook of Hydrology. McGraw-Hill, New York.
Strupczewski, W., Kochanek, K., Bogdanowicz, E., Markiewicz, I., 2012. On seasonal approach to flood frequency modelling. Part I: two-component distribution revisited. Hydrol. Process. 26, 705–716.
Thompson, C., Croke, J., 2013. Geomorphic effects, flood power, and channel competence of a catastrophic flood in confined and unconfined reaches of the upper Lockyer valley, southeast Queensland, Australia. Geomorphology 197, 156–169.
Vergara-Temprado, J., Ban, N., Schär, C., 2021. Extreme sub-hourly precipitation intensities scale close to the Clausius-Clapeyron rate over Europe. Geophys. Res. Lett. 48, e2020GL089506 https://doi.org/10.1029/2020GL089506.
Volpi, E., Fiori, A., Grimaldi, S., Lombardo, F., Koutsoyiannis, D., 2019. Save hydrological observations! Return period estimation without data decimation. J. Hydrol. 571, 782–792.
Von Storch, H., Bürger, G., Schnur, R., von Storch, J.S., 1995. Principal oscillation patterns: a review. J. Clim. 8 (3), 377–400.
Walker, A.M., 1968. A note on the asymptotic distribution of sample quantiles. J. Roy. Stat. Soc. B 30 (3), 570–575.
WRA, 2009. Storm Rainfall and Flood Flow Analysis for Typhoon Morakot. Water Resources Agency, Ministry of Economic Affairs, Taiwan, ROC (in Chinese).
Xekalaki, E., 2014. On the distribution theory of over-dispersion. J. Statist. Distribut. Appl. 1, 19. https://doi.org/10.1186/s40488-014-0019-z.