

Rainfall frequency analysis using a mixed GEV distribution: a case study for annual maximum rainfalls in South Korea

Philyong Yoon · Tae-Woong Kim · Chulsang Yoo

Published online: 7 October 2012
© Springer-Verlag Berlin Heidelberg 2012

Abstract Extreme rainfalls in South Korea result mainly from convective storms and typhoon storms during the summer. A proper way for dealing with the extreme rainfalls in hydrologic design is to consider the statistical characteristics of the annual maximum rainfall from two different storms when determining design rainfalls. Therefore, this study introduced a mixed generalized extreme value (GEV) distribution to estimate the rainfall quantile for 57 gauge stations across South Korea and compared the rainfall quantiles with those from conventional rainfall frequency analysis using a single GEV distribution. Overall, these results show that the mixed GEV distribution allows probability behavior to be taken into account during rainfall frequency analysis through the process of parameter estimation. The resulting rainfall quantile estimates were found to be significantly smaller than those determined using a single GEV distribution. The difference of rainfall quantiles was found to be closely correlated with the occurrence probability of typhoon and the distribution parameters.

Keywords Extreme rainfall · Frequency analysis · Mixed distribution · Rainfall quantile

P. Yoon
Department of Civil and Environmental Engineering, Hanyang University, Seoul 133-791, Korea
e-mail: pyvic@daum.net

T.-W. Kim (✉)
Department of Civil and Environmental Engineering, Hanyang University, Ansan 426-791, Korea
e-mail: twkim72@hanyang.ac.kr

C. Yoo
School of Civil, Environmental and Architectural Engineering, Korea University, Seoul 136-713, Korea
e-mail: envchul@korea.ac.kr

1 Introduction

Hydrologic frequency analysis is usually performed based on the appropriate probability distribution, which is selected on the basis of statistical tests for extreme hydrologic data collected in a specific region. In recent decades, however, various attempts have been made to address extreme hydrologic variables from a substantially different standpoint than that used in conventional frequency analysis. For example, Yue (2000, 2001), Yue and Rasmussen (2002), Zhang and Singh (2006), Kao and Govindaraju (2007) and Lee et al. (2010a) applied a bivariate distribution to hydrologic frequency analysis in order to address the joint probabilistic behavior between correlated variables. Strupczewski et al. (2001), Katz et al. (2002), Cunderlik and Burn (2003), Khaliq et al. (2006), Park et al. (2011), and Seo et al. (2012) proposed various methods to take the non-stationarity of hydrologic observations into consideration.

In addition to the issues mentioned above, one of most frequently issued in hydrologic frequency analysis is the problem of mixed distributions or multiple populations in hydrologic time series (Hirschboeck 1987). Homogeneity in the hydrologic time series is a basic underlying assumption for hydrologic frequency analysis. In South Korea, however, the annual distribution of daily rainfall has distinct two peaks. The first peak occurs during July that is attributed to stationary convective fronts such as the Changma. The second peak occurs during August that is associated with tropical storms such as typhoons (Lee et al. 2010b). The Changma provides more than 40 % of the annual rainfall, and the typhoons contribute about 25 % to the annual rainfall for the major river basins (Kim and Jain 2011). Consequently, Korean extreme rainfalls mainly result from the typhoon-induced storms and the convective storms during the monsoon season.

In a changing climate, since the interannual variability of convective storms is very larger and the contribution of typhoon storms to extreme rainfalls increases, the hydrologic risk due to extreme rainfall events will likely increase across South Korea. The practical engineering approach for dealing with extreme hydrologic events is to design flood protection structures based on the appropriate estimation of design rainfall and its corresponding design flood through hydrologic frequency analysis. Most of the hydrologic frequency analysis that is applied to practical situations involves extracting the annual maximum rainfalls for a specific rainfall duration using the observed rainfall data at a given location. The analysis then determines the best-fitting probability distribution function and calculates the design rainfall corresponding to the specific return period. The basic premise of this procedure is that the annual maximum rainfalls comprise a single population. However, as mentioned above, extreme rainfalls in South Korea are composed of rainfalls from convective storms and typhoons, which likely have different statistical characteristics. More specific discussions on this issue will be addressed in Sect. 2. Consequently, for an appropriate and effective hydrologic design, it is necessary to appropriately consider these two characteristics of extreme rainfalls in order to estimate the probability distributions by introducing mixed distributions.

Haan (1974) suggested the concept of analyzing zeros in hydrologic dataset in order to estimate a cumulative probability distribution. In addition, Haan (1974) suggested a mixed model, which combines a discrete distribution and a continuous distribution. Various kinds of mixed model have been applied to flood frequency analysis; Singh and Sinclair (1972) developed a mixed distribution model for an observed annual maximum daily flood from a heterogeneous flood population. Fiorentino et al. (1987) proposed a two component extreme value distribution to account for the characteristics of flood. Grego and Yates (2010) explored the use of finite mixture models in flood frequency analysis with annual maximum daily flows from two river gauges. Kedem et al. (1990) employed a mixed distribution model for determining an optimal choice of the threshold level of rainfall by combining a discrete and continuous distributions, such as log-normal, gamma, and inverse Gaussian distributions. A mixed model can be applied to investigate the effect of synoptic change on hydrologic extremes. For example, using a mixed gamma distribution, Yoo et al. (2005) investigated the effect of global warming on daily rainfalls.

This study focuses on applying a mixed distribution to rainfall frequency analysis in order to address the probability behavior of extreme rainfall in South Korea. To utilize a mixed distribution model in practice, maximum daily rainfalls must be classified as either typhoon rainfall or convective rainfall, after which the mixed GEV

distribution can be applied. A detailed description of the mixed model proposed in this study and its results are discussed in the following sections.

2 Extreme rainfall data

2.1 Rainfall data

This study collected daily rainfall data from 57 gauge stations across South Korea, as shown in Fig. 1. Note that rainfalls from convective storms are substantially different than those from typhoons in their physical and climatologic characteristics throughout the occurrence and development processes. However, since a convective storm and a typhoon interact in generating rainfall, it is technically difficult to clearly distinguish whether a measured rainfall from a site is caused by a convective storm or a typhoon. Hence, in order to utilize a mixed distribution, this study classified maximum daily rainfalls as either typhoon rainfalls or convective rainfalls. A typhoon rainfall was defined as rainfall measured while a typhoon classification was issued near the site; otherwise, the rainfall data were assumed to be from convective rainfall.

For example, Fig. 2 shows the daily rainfalls observed at the Seoul station from June 1 to September 30 in 2002. During this period, two typhoons affected Seoul, as shown in Fig. 3, and the maximum rainfall that occurred on July 5 became the annual maximum typhoon rainfall (AMTR) for 2002. These AMTRs were gathered every year to form the time series of the AMTR. Similarly, when typhoons were not issued, the maximum rainfall from August 7 became the annual maximum convective rainfall (AMCR) for 2002. These AMCRs were also gathered every year to form the time series of the AMCR. Then, the larger value between the AMTR and the AMCR became the annual maximum daily rainfall (AMDR). These AMDR values were similarly gathered every year to construct the time series of the AMDR. Figure 4 shows the relationship between the AMTR, AMCR, and AMDR. During the last 49 years (1961–2009), the number of years during which the AMTR became the AMDR is 7. Thus, the occurrence probability of typhoon is about 15 % for the Seoul station.

Figure 5 shows the scatter plots of the mean and variance for the AMTR, AMCR, and AMDR for all stations in Fig. 1. The mean values are observed in the order of AMDR, AMCR, and AMTR, while the variances are in order of AMTR, AMDR, and AMCR throughout South Korea. These plots indicate that the AMTR variability was the largest. In particular, the Gangneung station has a mean AMDR of 159.2 mm (the 9th largest station among 57 stations) and an AMDR variance of 14,473 mm² (the largest), giving the largest AMDR coefficient of variation

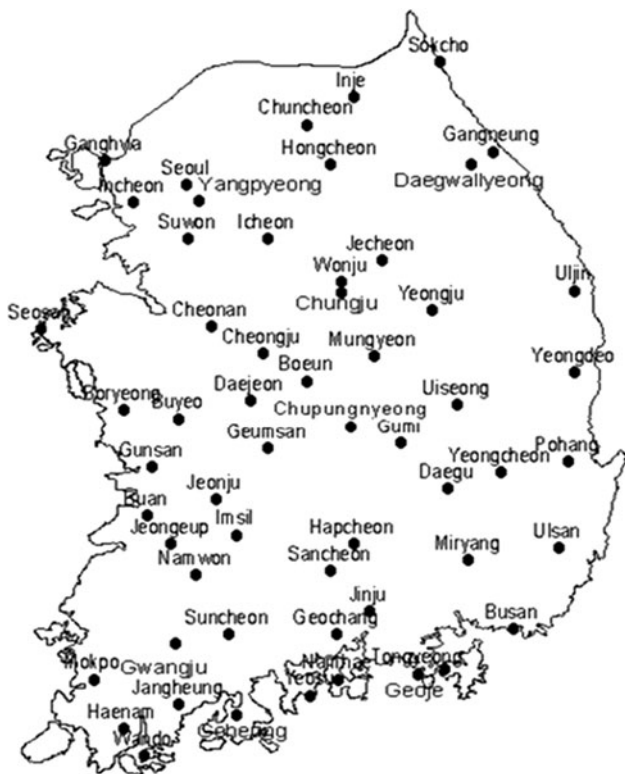


Fig. 1 Rainfall gauging stations used in this study. All stations maintain daily rainfalls measured more than 30 years

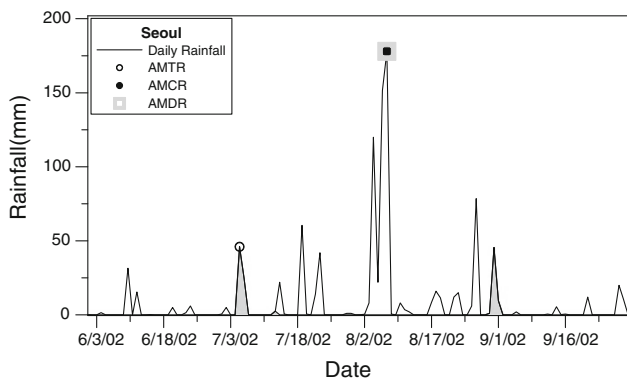


Fig. 2 Rainfalls observed during June and September in 2002. The AMTR (open circle) is the maximum rainfall during the period of typhoon issued (shaded duration), and the AMCR (closed circle) indicates the maximum rainfall among others. The bigger value between AMTR and AMCT become the AMDR

at 90.9 mm. At this same station, the mean AMCR was 111.2 mm (the 34th largest), the AMCR variance was 1,907 mm² (the 28th largest), and the AMCR coefficient of variation was 17.1 mm (the 28th largest). However, the mean AMTR was 130.3 mm (the 6th largest), the AMTR variance was 18,468 mm² (the largest), and the AMTR coefficient of variation was 141.8 mm (the largest).

Generally, a localized convective storm generates a high intensity of rainfall for a short period of time, whereas a

typhoon brings about heavy rainfall for relatively long period of time. In addition, there are distinct differences in the mean and variance between convective rainfalls and typhoon rainfalls, which ultimately result in different variabilities. Therefore, it is presumed that the magnitude and variability of extreme rainfalls at each station are related to the probabilistic characteristics of these two types of rainfall.

2.2 Hypothesis tests

Hypothesis testing is a common method of drawing inferences about a population based on statistical evidence from samples. Note that sample statistics, such as mean and variance, from the same population are generally statistically similar. In this study, the sample statistics from three different datasets—AMDR, AMCR, and AMTR, which were defined and constructed in Sect. 2.1—were identified by using hypothesis tests. In particular, a two-sample *t* test, a Wilcoxon rank-sum test, and a two-sample *F*-test were performed with a 5 % significance level to draw inferences about the mean, median, and variance, respectively.

A two-sample *t* test is to test whether the means are different and the test statistic is

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_{X1}^2}{n} + \frac{S_{X2}^2}{m}}} \tag{1}$$

where \bar{X}_1 and S_{X1}^2 are the mean and variance of the random variable *X* from population 1, respectively; \bar{X}_2 and S_{X2}^2 are the mean and variance of the random variable *X* from population 2, respectively; and *n* and *m* indicate the sample sizes of each population.

A Wilcoxon rank-sum test is also called a Mann–Whitney *U* test or Mann–Whitney–Wilcoxon test. This test is used to test whether two independent samples come from identical continuous distribution with equal medians. The Wilcoxon rank-sum test performs random testing by converting the values of variable into ranks. The test statistic is

$$z = \frac{k - \frac{nm}{2}}{\sqrt{\frac{nm(n+m+1)}{12}}} \tag{2}$$

where *k* is the rank sum.

A two-sample *F*-test is used to test whether two samples come from distributions with the same variance. The test statistic for a two-sample *F*-test is given in Eq. (3). The statistic has an *F*-distribution with *n*-1 and *m*-1 degree of freedom, if the null hypothesis of equality of variances is true.

$$F = \frac{S_{X1}^2}{S_{X2}^2} \tag{3}$$

In this study, the statistics of the two-sample *t* test, Wilcoxon rank-sum test, and two-sample *F*-test were

Fig. 3 The typhoons that were issued for South Korea during June–September in 2002. The typhoon Rammasun was categorized as three, issued on June 29 and dissipated on July 6. The typhoon Rusa was categorized as four, issued August 23 and dissipated on September 1

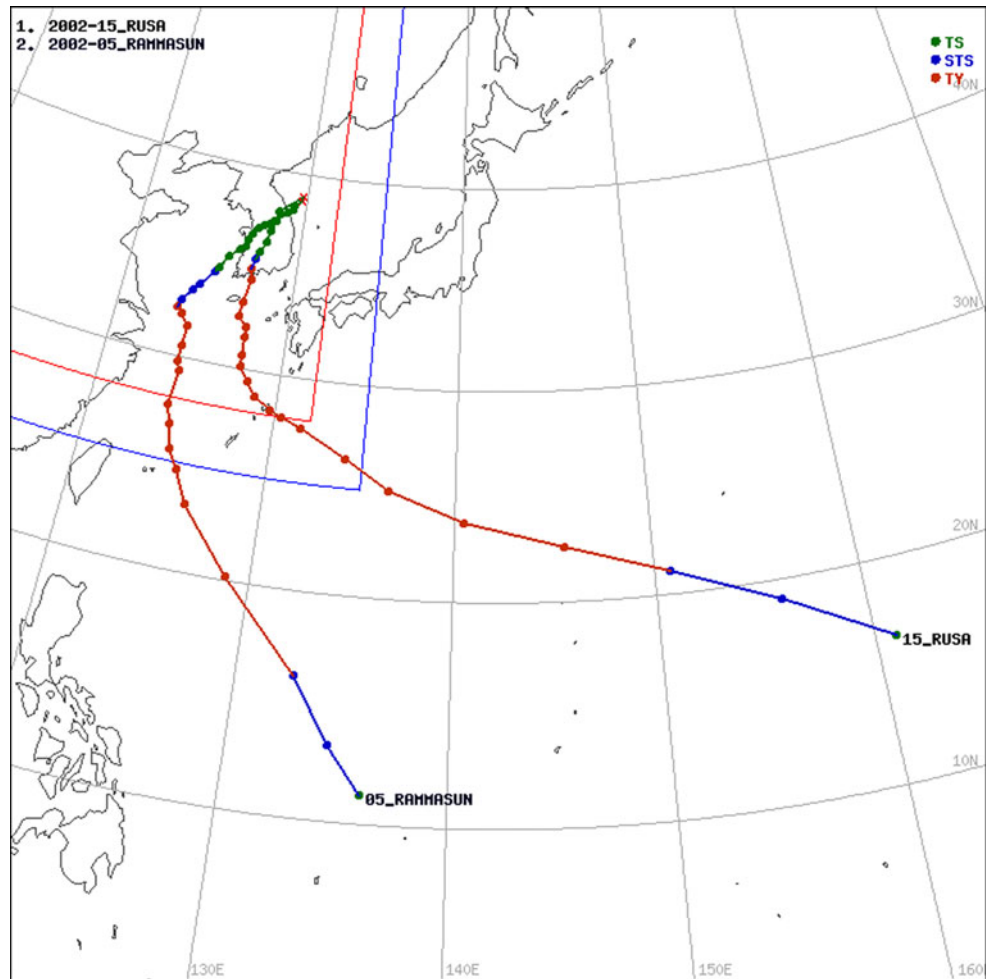
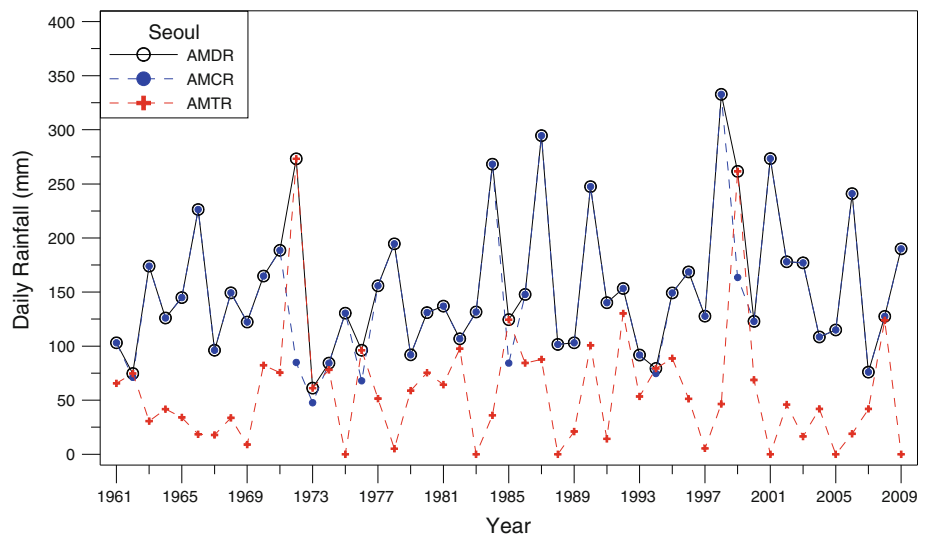


Fig. 4 The time series of AMTR, AMCR, and AMDR for the Seoul station



calculated using the Matlab Statistics Toolbox (Mathworks 2010). Based on the p value, the two-sample t test, Wilcoxon rank-sum test, and two-sample F-test verified the null hypothesis. The p value indicates the probability of the null hypothesis being true. In general, if the p value

is less than 0.05 at a significance level of 5 %, the null hypothesis is rejected. Table 1 provides the number of gauging stations where the null hypotheses were rejected by each testing method. More than 60 % of the stations rejected the null hypotheses for the AMCR and AMTR. In

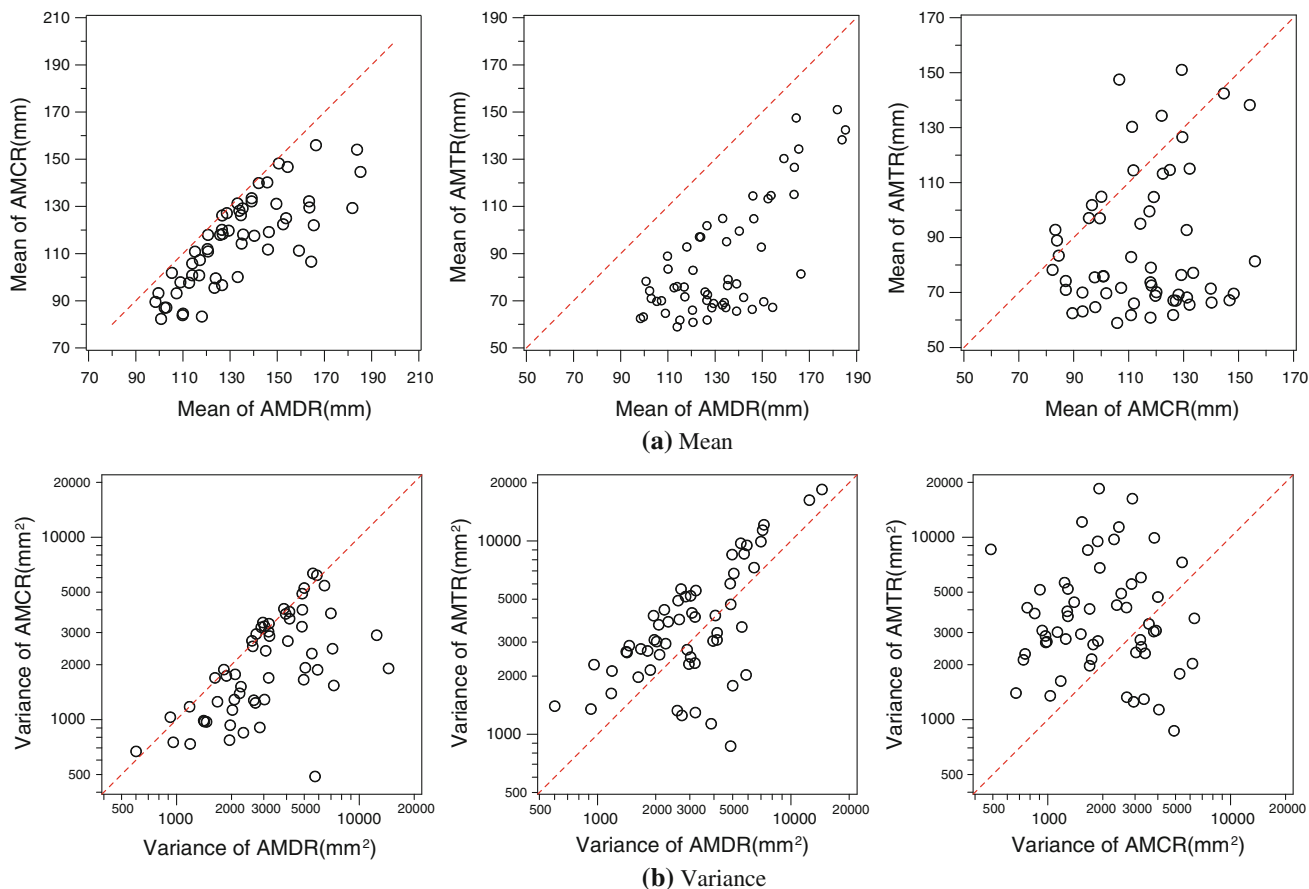


Fig. 5 Scatter plots of mean and variance among the AMTR, AMCR, and AMDR for the stations in Fig. 1

Table 1 The number of stations that reject the null hypotheses

	Tested sample series		
	AMDR vs AMCR	AMDR vs AMTR	AMCR vs AMTR
Two-sample <i>t</i> test	28 (49.1 %)	54 (94.7 %)	42 (73.7 %)
Wilcox rank sum test	22 (38.6 %)	52 (91.2 %)	36 (63.2 %)
Two-sample <i>F</i> -test	24 (42.1 %)	25 (43.9 %)	42 (73.7 %)

The null of hypotheses are that two samples have the same mean (two-sample *t* test), median (Wilcox rank sum test), and variance (two-sample *F*-test), respectively. The numbers in the table and the parentheses refer to the number of stations and the percentage of stations which reject the null hypothesis at 5 % significant level

particular, for the AMDR and AMTR, the two-sample *t* test and Wilcoxon rank-sum test were rejected at more than 90 % of the stations. These results imply that the AMCR and AMTR, which compose the AMDR, are datasets with statistically different characteristics. The conventional approach using a single distribution is not able to address the underlying probability behaviors responsible for each population in the parent population. Consequently, it is necessary to introduce a specific method that takes into account the statistical characteristics of each population.

3 Mixed GEV distribution

3.1 Probability distribution function

In general, a mixed distribution is formed by combining two different distributions. Accordingly, a mixed GEV distribution function is very suitable to analyze hydrologic phenomena in which specific events may occur because of two different causes (Haan 1974). For example, if $f_i(x), i = 1, 2, \dots, m$ is the probability density function of random variable *X* given that *X* is from the *i*th distribution, and the

parameter λ_i may be thought of as the probability that a random variable is from the probability density function $f_i(x)$ satisfying $\sum_{i=1}^m \lambda_i = 1$, then the probability density function and the cumulative distribution function of a mixed distribution would be expressed as in Eq. (4a) and (4b), respectively.

$$f_X(x) = \sum_{i=1}^m \lambda_i f_i(x) \tag{4a}$$

$$F_X(x) = \int_{-\infty}^{\infty} \sum_{i=1}^m \lambda_i f_i(t) dt \tag{4b}$$

Kedem et al. (1990) proposed the general form of a mixed distribution model, which has two different distribution functions, as shown in Eq. (5).

$$G(x) = p H_1(x) + (1 - p) H_2(x) \tag{5}$$

In this study, $H_1(x)$ and $H_2(x)$ represent the cumulative distribution functions for the typhoon rainfall and convective rainfall, respectively. Therefore, p indicates the percentage of typhoon rainfalls in the annual maximum rainfalls.

As investigated by Sultan et al. (2007) and based on Eq. (5), the probability density function (pdf) and the cumulative distribution function (cdf) of a mixed GEV model can be derived as in Eq. (6a) and (6b), respectively.

$$\begin{aligned} f(x) = & p \times \frac{1}{\alpha_1} \left[1 - \frac{\beta_1(x - x_1)}{\alpha_1} \right]^{\frac{1}{\beta_1} - 1} \\ & \times \exp \left[- \left(1 - \frac{\beta_1(x - x_1)}{\alpha_1} \right)^{\frac{1}{\beta_1}} \right] \\ & + (1 - p) \times \frac{1}{\alpha_2} \left[1 - \frac{\beta_2(x - x_2)}{\alpha_2} \right]^{\frac{1}{\beta_2} - 1} \\ & \times \exp \left[- \left(1 - \frac{\beta_2(x - x_2)}{\alpha_2} \right)^{\frac{1}{\beta_2}} \right] \end{aligned} \tag{6a}$$

$$\begin{aligned} F(x) = & p \times \exp \left[- \left(1 - \frac{\beta_1(x - x_1)}{\alpha_1} \right)^{\frac{1}{\beta_1}} \right] + (1 - p) \\ & \times \exp \left[- \left(1 - \frac{\beta_2(x - x_2)}{\alpha_2} \right)^{\frac{1}{\beta_2}} \right] \end{aligned} \tag{6b}$$

where α_1, β_1 , and x_1 are the parameters of scale, shape, and location for typhoon rainfall, while α_2, β_2 , and x_2 are the parameters of scale, shape, and location for convective rainfall. The cdf in Eq. (6b) is related to the non-exceedance probability; hence, the exceedance probability becomes $1 - F$.

In this study, the estimation of the rainfall quantiles corresponding to various return periods can be computed following Chow et al. (1988) such that

$$\frac{1}{T} = p(x \geq x_T) = 1 - F(x_T) \tag{7a}$$

$$x_T = F^{-1} \left(1 - \frac{1}{T} \right) \tag{7b}$$

where x_T is the rainfall quantile corresponding to return period T , and F^{-1} represents the inverse function of F in Eq. (6b)

3.2 Parameter estimation

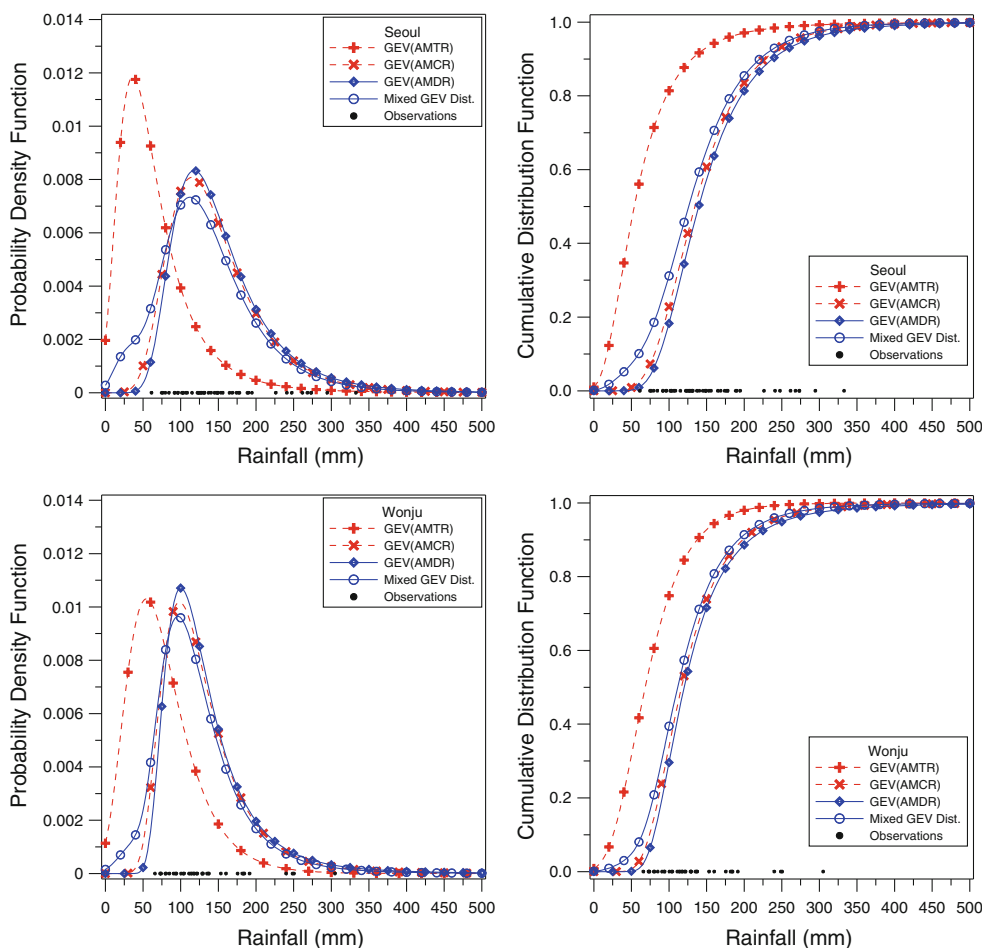
Equation (6) presents seven parameters for the mixed GEV model ($p, \alpha_1, \beta_1, x_1, \alpha_2, \beta_2, x_2$). In order to estimate p , the number of AMTRs in the AMDR time series was counted, and is presented in Fig. 4. For example, the number of AMTRs to become the AMDR between the year 1961 and the year 2009 was 7 for the Seoul station, giving $p = 0.1429$. Similarly, the value for p was calculated for all the stations. The resulting mean value of p was 0.2830, the minimum was 0.0435 (at Suwon), and the maximum was 0.6757 (at Sancheong).

Since the numbers of sample from each population (e.g., AMTR and AMCR) are very small in the parent population (e.g., AMDR), moreover, the distribution parameters should represent the statistical characteristics of each population in Eq. (6), a special approach will be required to estimate proper parameter estimates. This study applied the method of maximum likelihood (ML) for the AMTR data in order to estimate α_1, β_1 , and x_1 , and likewise for the AMCR data in order to estimate α_2, β_2 , and x_2 . Note that Yoo et al. (2005) similarly estimated distribution parameters from two different populations that composed a single dataset for daily rainfall. Figure 6 shows graphical representations of the pdf and cdf for various rainfall datasets used in this study. The figures confirmed that the mixed model satisfied the constrain of a probability function, such as $\int f(x) dx = 1$. The important role of individual typhoon events and their contributions to the design rainfalls are evident from analyses presented in Fig. 6. It is important to note that the typhoon rainfall shifts the probability distribution of extreme rainfall to the left.

3.3 Goodness-of-fit test

In this study, a χ^2 test and a Kolmogorov–Smirnov (K–S) test were performed to determine the goodness of fit for the proposed mixed distribution function. The test statistic of the χ^2 test is given by Eq. (8), after calculating the sample frequency and the theoretical frequency for the specific probability distribution function.

Fig. 6 Comparisons of probability density functions and cumulative distribution functions for the Seoul and Wonju stations



$$x^2 = \sum_{i=1}^k \left(\frac{n_i - np_i}{np_i} \right)^2 \tag{8}$$

where n is the number of sample data points and k is the number of class intervals. Therefore, n_i and np_i are the frequencies and theoretical frequencies in the i th interval of sample data, respectively, and p_i is the theoretical probability for the i th interval of data. The value of χ^2 approaches a χ^2 distribution with the degree of freedom $\nu = k - h - 1$, where h is the number of parameters. In this study, the χ^2 test was performed at a significance level of 5 % ($\alpha = 5\%$, $\chi_{0.95,5}^2 = 11.1$).

The test statistic of the K–S test is the maximum deviation between the empirical distribution and the theoretical distribution, which is calculated by Eq. (9).

$$D_{\text{Max}} = \text{Max}|F(x) - F_0(x)| \tag{9}$$

where $F(x)$ the empirical distribution of the observed data, and $F_0(x)$ is the theoretical distribution. If the value of D_{Max} is larger than the critical value D_n^z , then the null hypothesis is rejected. This test was performed in this study

with a significance level of 5 % ($D_n^z = 0.1943$). The test results show that the proposed GEV distribution could be applied to most of the stations considered in this study.

4 Applications and discussions

4.1 Rainfall frequency curves

Quantile estimations of daily maximum rainfalls corresponding to various return periods were calculated for all gauging stations by inverting the exceedance probability, as described in Sect. 3.

Figure 7 shows the rainfall frequency curves comparing the rainfall quantiles for Hongcheon and Tongyoung stations as representative cases. Rainfall frequency curves were derived from each annual maximum time series, which all had the same record period. The results show a much more varied pattern of change in extreme rainfall frequencies across South Korea than when a single GEV distribution being applied.

Fig. 7 Comparisons of rainfall quantiles corresponding to return periods for **a** Hongcheon and **b** Tongyeong station

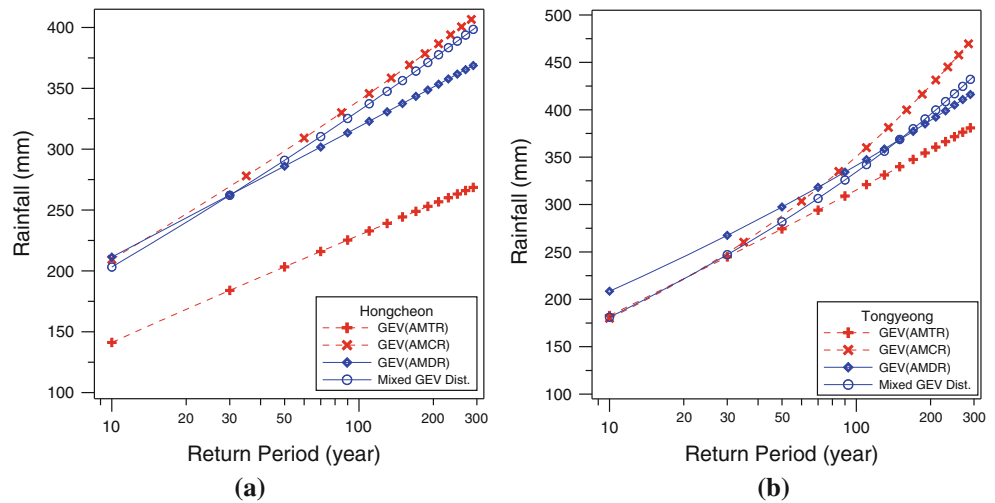
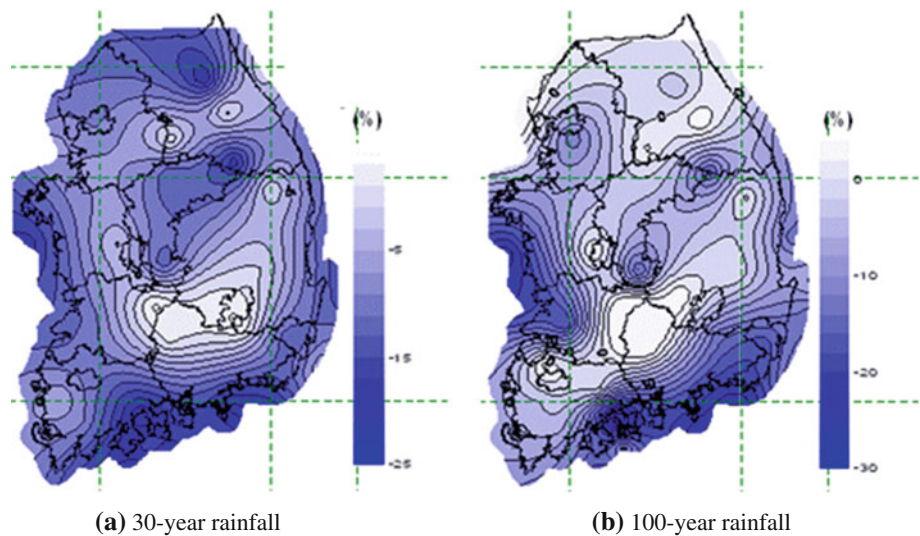


Fig. 8 Ratio of decrease in rainfall quantiles



Generally, using the GEV distribution, the AMCR rainfall quantile becomes larger than that for the AMTR, while rainfall quantiles estimated by the mixed GEV model are smaller than rainfall quantiles (with GEV) for the AMDR. However, there were cases of inverted frequency curves for each site. For example, inverted frequency curves were observed in 9 stations (Daejeon, Yeosu, Ganghwa, Yangpyeong, Inje, Jecheon, Hongcheon, Boeun, and Geumsan) for small return periods, and the p values from these stations were relatively large in comparison to those from other sites. In contrast, inverted frequency curves were observed in 13 stations (Incheon, Pohang, Gunsan, Daegu, Jeonju, Busan, Tongyeong, Mokpo, Jinju, Cheonan, Mungyeong, Youngcheon, and Milyang) for large return periods with relatively small p values when compared with those from other sites.

The results are summarized on the map in Fig. 8 in terms of percentage change (i.e.,

$$PC_T = 100 \times \frac{P_{T,MIX} - P_{T,GEV}}{P_{T,GEV}}$$

where $P_{T,MIX}$ refers to the rainfall quantiles estimated from the mixed GEV model, and $P_{T,GEV}$ indicates the rainfall quantiles acquired from the GEV distribution) for the 30-year and 100-year rainfalls. The 30-year rainfall was applied mainly as a design criterion for small-sized flood structures, such as culvert or urban rainwater drainage systems, while the 100-year rainfall was applied as a design standard for medium to large scale flood structures, such as spillways and river banks. The frequency of occurrence of flood damage was formally set to an average of once in 100 years (the so-called 100-year flood) in the Flood Disaster and Protection Act of 1973. However, the 100-year flood had been used in engineering design for many years before 1973 (WMO 2006).

The estimated rainfall quantiles showed decreasing trends throughout South Korea, in general. These

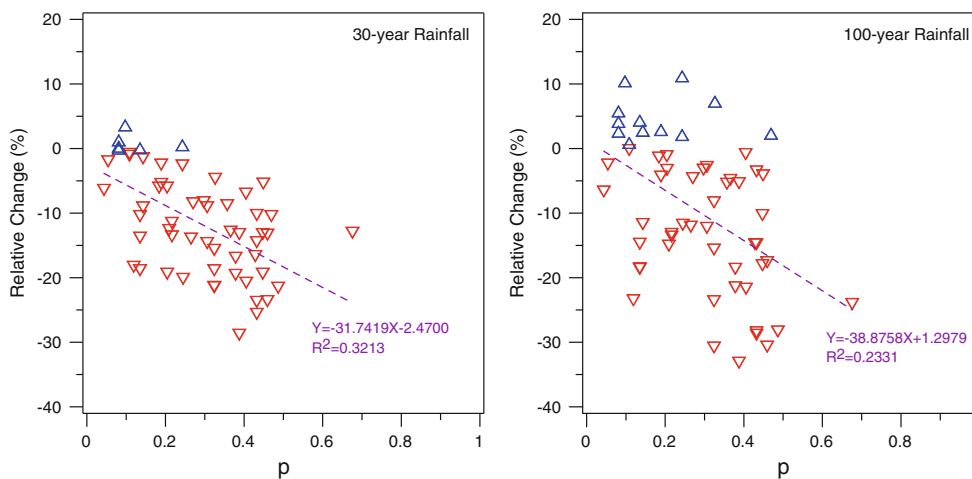


Fig. 9 Relations between relative change and p

decreasing trends were especially prominent in some parts of the western and southern regions. Overall, most of stations showed a substantial decrease in extreme rainfalls for both the 30-year and the 100-year return periods. Park et al. (2011) estimated the 100-year rainfall using the annual maximum of daily rainfall and showed that eastern and south-western areas have high return values for the rainfall quantile. Incorporating the results of Park et al. (2011), the eastern region remains high risk for extreme rainfalls while western and southern regions may reduce return levels. The decreased extreme rainfalls in these regions are certain to contribute to a decreased flood risk. In addition, the small contribution of typhoon rainfall to the 30-year and the 100-year rainfall results from the spatial distribution of typhoon-induced precipitation which is influenced by topography (Kim and Jain 2011). This factor indicates that frequency analysis using a single distribution function conventionally applied in practice may overestimate extreme rainfalls and floods. A particular utility of the analysis approach utilized herein is that the relative role of typhoon and convective rainfalls can be analyzed to understand the spatial patterns and the causal factors contributing to the observed trends in the frequency of heavy rainfalls. The issue is discussed in detail in Sect. 4.2.

4.2 Relations between changes in rainfall quantiles and distribution parameters

An improved understanding of the atmospheric and oceanic controls on non-typhoon precipitation provides useful insights regarding the mechanisms that underpin warm season hydroclimatology and the future trends over the Korean peninsula (Kim and Jain 2011).

In this section, the relationship between the relative percentage change (PC) values in Fig. 8 and the distribution parameters are examined. First, as shown in Fig. 9,

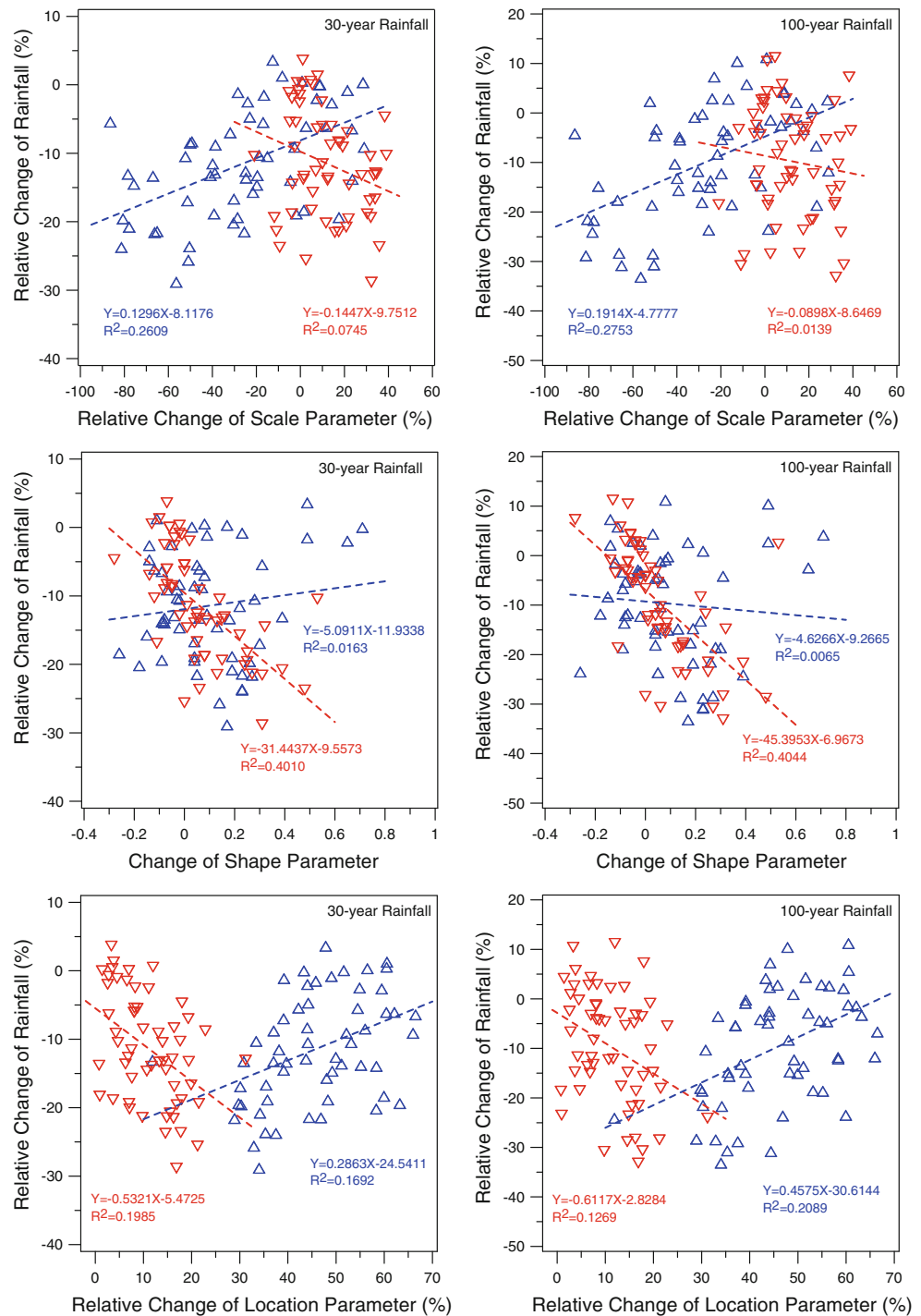
larger p values correspond to smaller relative variations, which imply rainfall quantiles estimated through the mixed GEV model become smaller relative to those from a GEV function. However, in case of the 30-year rainfall, the rainfall quantiles by the mixed GEV model from 6 stations (Boeun, Hongcheon, Yangpyeong, Ganghwa, Geumsan, and Jeonju) were larger than those by the mixed GEV model, while for the 100-year rainfall, the rainfall quantiles by the mixed GEV model were larger than those from 12 stations (Jecheon, Inje, Pohang, Yangpyeong, Gunsan, Cheonan, Boeun, Hongcheon, Yeosu, Daejeon, and Ganghwa).

The PC value of the rainfall quantile shows a close relationship not only with the probability of typhoon rainfall p , but also with the distribution parameters. As shown in Fig. 10, the change in the distribution parameters for the AMTR ($\frac{\alpha_0 - \alpha_2}{\alpha_0}, \beta_0 - \beta_2$, and $\frac{x_0 - x_2}{x_0}$) and in the relative PC are positively correlated, whereas the change in the distribution parameters for the AMCR ($\frac{\alpha_0 - \alpha_1}{\alpha_0}, \beta_0 - \beta_1$, and $\frac{x_0 - x_1}{x_0}$) and in the relative PC are negatively correlated. Note that α_0, β_0 , and x_0 are the parameters of scale, shape, and location for the AMDR respectively. The largest coefficient of determination for relative percentage changes was found in the change of the shape parameter for the AMCR. For both the 30- and 100-year rainfalls, the R^2 values were about 0.4.

5 Conclusions

Extreme rainfalls from a convective storm have different statistics, especially temporal variability, with those from a typhoon. When we need to determine design rainfalls in the assessment of flood risk, it is necessary to distinguish the type of extreme rainfall. In this study, rainfall quantiles

Fig. 10 Relations between relative change of rainfall and relative change of distribution parameters. *up-pointing triangle* indicate the relationship between the distribution parameters of the AMDR and the AMTR, and *down-pointing triangle* shows the relationship between the distribution parameters of the AMDR and the AMCR



were calculated for the annual maximum daily rainfall by applying a mixed distribution function, and the relations between parameters of the mixed distribution function were examined. The annual maximum rainfall (AMDR) was formed using the typhoon rainfall (AMTR) and the convective rainfall (AMCR) according to the type of storm, and the statistical characteristics of each rainfall time series were different. Accordingly, the distribution function

model was implemented so that statistical probability characteristics of the AMTR and AMCR could be reflected appropriately. The rainfall quantiles estimated by the mixed distribution model were compared with the rainfall quantiles from the conventional method, which directly applied a single GEV function to the AMDR. Although most of the rainfalls that constituted the AMDR were extracted from the AMCR, and typhoon occurrence was

irregular with a low probability, its presence was found to have a large effect on the rainfall quantiles.

The application of the mixed GEV model suggested by this study showed that the rainfall quantiles tended to decrease overall when compared to those cases using the GEV function on annual maximum time series, as is normally the case. This discrepancy occurred because the distribution characteristics of typhoon rainfalls are largely different from those of convective rainfalls as well as those from annual maximum rainfalls. More specifically, most of the AMTR distribution was biased toward the left side, which reduced the rainfall quantile. Most notably, the overall results indicate that the use of a mixed distribution leads to a more feasible and economic design criteria for hydrologic design.

Although the proposed mixed model was applied fairly well to the daily maximum rainfall in South Korea, a number of issues may still remain to be clarified. For example, this study classified extreme rainfalls into the convective rainfall and the typhoon rainfall, based on whether a typhoon was issued or not. Although it is technically difficult to clearly distinguish extreme rainfalls based on their mechanisms at this moment, further works contribute to suggesting a new concept for the decomposition of rainfall data based on physical grounds.

With the mixed distribution function model for the estimation of rainfall quantile proposed in this study, the effect of an extreme rainfall event or a typhoon caused by repeated climate changes can be considered, and utilized as useful basic data in preparation of a design quantile standard or rainfall quantile maps which are renewed every 10–20 years. Thus, these results can be used in designing more efficient flood protection structures.

Acknowledgments This work was supported by grants from the National Research Foundation (No. 2010-0015578), Ministry of Education, Science and Technology, and Natural Hazard Mitigation Research Group (NEMA-NH-2011-42), National Emergency Management Agency, South Korea. The authors also thank the anonymous reviewers for their constructive comments and corrections.

References

- Chow VT, Maidment DR, Mays LW (1988) Applied hydrology. McGraw-Hill, New York
- Cunderlik JM, Burn DH (2003) Non-stationary pooled flood frequency analysis. *J Hydrol* 276:210–223
- Fiorentino M, Arora K, Singh VA (1987) The two-component extreme value distribution for flood frequency analysis: derivation of a new estimation method. *Stoch Hydrol Hydraul* 1:199–208
- Grego JM, Yates PA (2010) Point and standard error estimation for quantiles of mixed flood distributions. *J Hydrol* 391:289–301
- Haan CT (1974) Statistical methods in hydrology. The Iowa State University Press, Ames
- Hirschboeck KK (1987) Hydroclimatologically-defined mixed distributions in partial duration flood series. In: Singh VP (ed) Hydrologic frequency modeling. D. Reidel Publishing Company, Dordrecht, pp 230–257
- Kao S-C, Govindaraju RS (2007) A bivariate frequency analysis of extreme rainfall with implications for design. *J Geophys Res* 112:D13119. doi:10.1029/2007JD008522
- Katz RW, Parlange MB, Naveau P (2002) Statistics of extremes in hydrology. *Adv Water Resour* 25:1287–1304
- Kedem B, Chiu LS, Karni Z (1990) An analysis of the threshold method for measuring area-average rainfall. *J Appl Meteorol* 29:3–20
- Khaliq MN, Ouarda TBMJ, Ondo J-C, Gachon P, Bobée B (2006) Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: a review. *J Hydrol* 329:534–552
- Kim J-S, Jain S (2011) Precipitation trends over the Korean peninsula: typhoon-induced changes and a typology for characterizing climate-related risk. *Environ Res Lett* 6:034033. doi:10.1088/1748-9326/6/3/034033
- Lee CH, Kim T-W, Chung G, Choi M, Yoo C (2010a) Application of bivariate frequency analysis to the derivation of rainfall-frequency curves. *Stoch Environ Res Risk Assess* 24:389–397
- Lee S-S, Vinayachandran PN, Ha K-J, Jhun J-G (2010b) Shift of peak in summer monsoon rainfall over Korea and its association with El Niño-Southern Oscillation. *J Geophys Res* 115:D02111. doi:10.1029/2009/2009JD011717
- Mathworks (2010) Statistics toolbox; user's guide. The Mathworks Inc., Natick
- Park J-S, Kang H-S, Lee YS, Lim M-K (2011) Changes in the extreme daily rainfall in South Korea. *Int J Climatol* 31:2290–2299. doi:10.1002/joc.2236
- Seo L, Kim T-W, Choi M, Kwon H-H (2012) Constructing rainfall depth-frequency curves considering a linear trend in rainfall observations. *Stoch Environ Res Risk Assess* 26:419–427
- Singh VP, Sinclair (1972) Two-distribution method for flood frequency analysis. *Water Resour Res* 7:1144–1150
- Strupczewski WG, Singh VP, Mitosek HT (2001) Non-stationary approach to at-site flood frequency modeling. III. Flood analysis of Polish rivers. *J Hydrol* 248:152–167
- Sultan KS, Ismail MA, Al-Moisheer AS (2007) Mixture of two inverse Weibull distributions: properties and estimation. *Comput Stat Data Anal* 51:5377–5387
- WMO (2006) Comprehensive risk assessment for natural hazards. WMO/TD No. 955
- Yoo C, Jung K-S, Kim T-W (2005) Rainfall frequency analysis using a mixed Gamma distribution: evaluation of the global warming effect on daily rainfall. *Hydrol Process* 19:3851–3861
- Yue S (2000) The Gumbel mixed model applied to storm frequency analysis. *Water Resour Manag* 14:377–389
- Yue S (2001) A bivariate gamma distribution for use in multivariate flood frequency analysis. *Hydrol Process* 15:1033–1045
- Yue S, Rasmussen P (2002) Bivariate frequency analysis: discussion of some useful concepts in hydrological application. *Hydrol Process* 16:2881–2898
- Zhang L, Singh VP (2006) Bivariate flood frequency analysis using the copula method. *J Hydrol Eng* 11(2):150–164