# Comparison of annual maximum rainfall events of modern rain gauge data (1961–2010) and Chukwooki data (1777–1910) in Seoul, Korea

Chulsang Yoo, Minkyu Park, Hyeon Jun Kim and Changhyun Jun


## ABSTRACT

In this study, the annual maximum rainfall event series were constructed and compared for both the modern flip-bucket type rainfall data, collected since 1961 (the modern data), and the old Chukwooki rainfall data, collected from 1777 to 1910 (the Chukwooki data). First, independent rainfall events were derived, by applying the same rainfall threshold of 2 mm and data collection time interval of 2 hours, to both the Chukwooki and the modern data. Annual maximum rainfall event series were then constructed, by applying Freund's bivariate exponential distribution annually. Finally, bivariate frequency analysis was done for the annual maximum rainfall event series constructed, by applying the bivariate logistic model to evaluate and quantify their characteristics. The results are in summary: (1) characteristics of the Chukwooki rainfall events and modern rainfall events are very similar to each other; (2) the annual maximum rainfall events of modern data are slightly larger than those of the Chukwooki data. The total rainfall depth per rainfall event for any given return period is thus estimated to be a little higher for the modern data than that of the Chukwooki data. However, based on the findings in this study, it could not be concluded that the rainfall characteristics have significantly changed during the last 200 years.


**Key words** | annual maxima, bivariate frequency analysis, Chukwooki, rainfall event


**Chulsang Yoo** (corresponding author)
Department of Civil, Environmental and Architectural Engineering, College of Engineering,
Korea University,
Seoul 136-705,
Korea
E-mail: *envchul@korea.ac.kr*

**Minkyu Park**
Department of Disaster Mitigation and Safety Science, Faculty of Convergence Science,
Jungwon University,
Goesan-gun, Chungcheongbuk-do 367-805,
Korea

**Hyeon Jun Kim**
Korea Institute of Civil Engineering and Building Technology,
University of Science and Technology,
Goyang-si 411-712,
Korea

**Changhyun Jun**
School of Civil, Environmental and Architectural Engineering, College of Engineering,
Korea University,
Seoul 136-713,
Korea


## INTRODUCTION

It is generally held in climate change research that long data length has a special meaning (Stedinger & Cohn 1986; Gergis & Fowler 2005; Wang *et al.* 2006; Yoo 2006; Yoo *et al.* 2007). The results derived can be more significant with longer data length. A possible fluctuation of rainfall characteristics can also easily be detected with longer data length (Frei & Schär 2001; Yoo 2008). Unfortunately, however, the length of systematic rainfall measurements is very limited in most countries, including Korea.

In Korea, systematic modern rainfall measurements have been available since 1960. Before then, the daily rainfall depth, along with sporadic 3-hour and 6-hour rainfall, has been recorded since 1907 (KMA 2004). Before 1907, an old Korean rain gauge, named Chukwooki, was used to measure rainfall in the major cities in Korea (Figure 1). In fact, the Chukwooki was invented in 1441, and the longest data available is in Seoul, since 1777. The data structure of the Chukwooki rainfall is very basic, with simply the starting time, ending time, and the total rainfall depth of a rainfall event. That is, only the duration and total rainfall depth of a rainfall event were recorded (Jhun & Moon 1997; Kim *et al.* 2007). Fortunately, the old rainfall data that are available have been proven to have sufficient quality for further analysis, along with modern rainfall data (Jung 1999; Wang *et al.* 2006; Yoo 2006).

doi: 10.2166/wcc.2017.110


**Figure 1** | Chukwooki, an old Korean rain gauge used in the Choson Dynasty.

By simply including all the rainfall data available, the total length of rainfall data in Seoul becomes more than 250 years. This could present valuable information for research on climate change or frequency analysis. However, even the quantification of rainfall itself is very tricky, as the modern and Chukwooki data structures are totally different from one another. As mentioned above, the data structure of Chukwooki rainfall is bivariate, providing the rainfall duration and total rainfall depth of a rainfall event. As conventional univariate frequency analysis could not handle these rainfall data, it is quite natural to explore the possible use of bivariate analysis, targeting the rainfall event itself (Yue *et al.* 1999; Yue & Wang 2004; Park *et al.* 2014).

The main objective of this study is to construct and compare the annual maximum rainfall event series for both the modern flip-bucket type and the old Chukwooki rainfall data. Bivariate frequency analysis is applied to the constructed annual maximum rainfall event series, to evaluate and quantify their characteristics. The results derived are finally used to compare the rainfall events in the Choson Dynasty about 200 years ago with those nowadays. The results can also be interpreted with regard to climate change due to global warming.

## BIVARIATE FREQUENCY ANALYSIS OF RAINFALL EVENTS

### Bivariate distributions

Various bivariate distributions have been used for frequency analysis in hydrology. The Gumbel bivariate exponential distribution was used to model the rainfall intensity and rainfall duration of extreme rainfall events (Bacchi *et al.* 1994). For the frequency analysis of rainfall events, the bivariate Gamma distribution (Yue 2001b) and the bivariate extreme value distribution (Tawn 1988, 1990; Coles & Tawn 1991) were also considered. The bivariate logistic and mixed models, with Gumbel margins, were also popular in the analysis of rainfall events and floods (Yue *et al.* 1999; Yue 2001a; Yue & Wang 2004). In this study, following the procedure of bivariate frequency analysis of rainfall events (Park *et al.* 2014), the bivariate exponential distribution was used for the construction of annual maximum rainfall event series and then the constructed annual maximum rainfall event series were analyzed by the bivariate logistic model.

In general, a bivariate distribution has its application limit due to the given constraints of the parameters or the given range of correlation coefficients. The Freund's bivariate exponential mixture distribution considered in this study has a rather wide range of correlation coefficients for its application, from $-0.333$ to 1.0. As the analytical solution of the maximum likelihood estimator is available, the model parameters can also easily be estimated. Freund's bivariate exponential distribution is given by the following joint probability distribution function (pdf) (Kotz *et al.* 2000):

$$\begin{aligned} P_{X_1,X_2}(x_1, x_2) &= \alpha_1 \alpha_2' e^{-\alpha_2' x_2 - \gamma_2 x_1}, \quad 0 \le x_1 < x_2 \\ &= \alpha_1' \alpha_2 e^{-\alpha_1' x_1 - \gamma_1 x_2} \quad\ 0 \le x_2 < x_1 \end{aligned} \tag{1}$$

where parameters $\alpha_1 > 0$, $\alpha_2 > 0$, $\alpha_1' > 0$, $\alpha_2' > 0$, and $\gamma_i = \alpha_1 + \alpha_2 - \alpha_i'$ $(i = 1, 2)$. The two variables, $X_1$ and $X_2$, are assumed to follow exponential distributions, with parameters $\alpha_1$ and $\alpha_2$, respectively. The dependence between $X_1$ and $X_2$ also changes the parameters of the distribution. In particular, when $X_1$ and $X_2$ are independent, $\alpha_1 = \alpha_1'$ and $\alpha_2 = \alpha_2'$. When $X_1$ and $X_2$ are dependent, the parameters $\alpha_1$, $\alpha_2$, $\alpha_1'$ and $\alpha_2'$ are generally estimated by the method of maximum likelihood (Freund 1961):

$$\hat{\alpha}_1 = \frac{r}{\sum^* X_1 + \sum^{**} X_2} \tag{2}$$

$$\hat{\alpha}_2 = \frac{n-r}{\sum^* X_1 + \sum^{**} X_2} \tag{3}$$

$$\hat{\alpha}_1' = \frac{n-r}{\sum^* X_1 + \sum^{**} X_2} \tag{4}$$

$$\hat{\alpha}_2' = \frac{r}{-\sum^{**} X_1 + \sum^* X_2} \tag{5}$$

where $r$ is the number of rainfall events in a random sample of size $n$, $X_1'$ the standardized variable of $X_1$, defined as $(X_1 - \mu_{X_1})/\sigma_{X_1}$, using its location parameter, $\mu_{X_1}$, and scale parameter, $\sigma_{X_1}$, where $X_2'$ is similarly $(X_2 - \mu_{X_2})/\sigma_{X_2}$. $\sum^* X_1$ denotes the sum of the variables $X_1$ when $X_1'$ is greater than $X_2'$, and $\sum^{**} X_1$ the sum of the variables $X_1$ when $X_2'$ is greater than $X_1'$. Similarly, $\sum^* X_2$ denotes the sum of the variables $X_2$ when $X_1'$ is greater than $X_2'$, and $\sum^{**} X_2$ the sum of the variables $X_2$ when $X_2'$ is greater than $X_1'$.

Three bivariate extreme value distributions are popular: the mixed model, the logistic model and the biextremal model (Kotz *et al.* 2000). Among these three models, the logistic model was considered in this study. The logistic model was originally proposed by Gumbel (1960), which has been evaluated as being very flexible without any serious limitations (Serinaldi & Grimaldi 2007). This model has also been applied to many hydrologic problems (Coles 1993; Yue 2001a, 2001c; Yue & Wang 2004; De Michele *et al.* 2005; Smith 2005; Vannitsem & Naveau 2007; Muller *et al.* 2008).

The probability distribution function of the bivariate logistic model, or formally the bivariate extreme distribution of logistic model with the standard generalized extreme value (GEV) marginal distributions, is given as (Gumbel 1960, 1965):

$$F_{Y_1,Y_2}(y_1, y_2) = \exp\left[-(y_1^m + y_2^m)^{(1/m)}\right],\ m \geq 1 \tag{6}$$

where $m$ is the parameter describing the association between two variables of $Y_i$ $(i = 1, 2)$. Also, $Y_i$ is the transformed variable of the original variables $X_i$, based on the following Fréchet transformation:

$$Y_i = \left[1 + S_i \frac{X_i - a_i}{b_i}\right]^{-(1/S_i)} \tag{7}$$

where $a_i$, $b_i$ and $S_i$ are the location, scale and shape parameters of $X_i$, respectively. Under the assumption that $X_1 \cdots X_n$ are independent variables, having a GEV distribution, the log-likelihood function for the GEV parameters is provided as follows (Kotz *et al.* 2000; Rao & Hamed 2000):

$$L(a_i, b_i, s_i) = -n \log(a_i) - \left(1 + \frac{1}{s_i}\right) \sum_{i=1}^{n} \log\left[1 + s_i \frac{X_i - a_i}{b_i}\right] - \sum_{i=1}^{n} \left[1 + s_i \frac{X_i - a_i}{b_i}\right]^{-(1/s_i)} \tag{8}$$

where $1 + s_i(X_i - a_i)/b_i > 0$ $(i = 1, \cdots, n)$.

Maximization of Equation (8) leads to the maximum likelihood estimates of the location, scale and shape parameters of the marginal variables. There is no analytical solution, but for any given dataset, the maximization can be easily achieved using standard numerical optimization algorithms (Kotz *et al.* 2000; Rao & Hamed 2000). Additionally, in order to estimate the parameter, $m$, from a sample of size, $n$, one can use the observed frequencies in the $2 \times 2$ table formed by dichotomizing each variable at its sample median (Kotz *et al.* 2000). The resulting estimate, $\hat{m}$, is expressed as:

$$\hat{m} = \left\{\log\left(-\frac{\log \hat{p}}{\log 2}\right)\right\}^{-1} \log 2 \tag{9}$$

where $\hat{p}$ indicates the observed proportion of rainfall events, i.e. the number of rainfall events divided by $n$.

### Bivariate frequency analysis

A distribution-free multivariate Kolmogorov-Smirnov goodness-of-fit test (Justel *et al.* 1997) was applied to test the validity of the bivariate distributions considered in this study. This test evaluates the difference between empirical and theoretical joint cumulative probabilities of real events. If no difference at a significance level $\alpha$ is detected, the bivariate distribution is assumed suitable for representing the data. The theoretical joint cumulative probabilities can be calculated by integrating Equation (1) when the bivariate exponential distribution is considered or Equation (7) when the bivariate logistic model is considered. The empirical joint cumulative probabilities can also be estimated using a formula for the plotting position. In this study, the following Weibull formula was extended to the bivariate case. (Yue 2001a; Lee *et al.* 2010):

$$P_k = \frac{k}{N+1} \tag{10}$$

where $P_k$ is the empirical cumulative probability, for which a given value is less than the $k$-th smallest observation in the dataset of $N$ observations.

Several different definitions of bivariate return periods exist in the frequency analysis of bivariate events (Shiau 2003; Salvadori & De Michele 2007). Of these, two different definitions of return periods are general. The first is the so-called 'OR return period' (Yue & Rasmussen 2002; De Michele *et al.* 2005). For a fixed return period T, the occurrence of one joint event may be sufficient for either $X_1$ or $X_2$ to exceed the given thresholds ($X_1 > x_1$ or $X_2 > x_2$). This joint event is called the 'OR joint event', and the related return period is called the 'OR joint return period'. The second is the case where the joint event needs to happen when both $X_1$ and $X_2$ are larger than the prescribed values ($X_1 > x_1$ and $X_2 > x_2$). This joint event is called the 'AND joint event', and the related return period is called the 'AND joint return period'. Between these two, this study considered the AND return period, as it was assumed to be more suitable to the hydrologic frequency analysis (Park *et al.* 2014). The AND joint return period, $T^*(x_1, x_2)$, is defined as follows:

$$T^*(x_1,x_2) = \frac{1}{F^*(x_1,x_2)} = \frac{1}{1+F(x_1,x_2)-F_{X_1}(x_1)-F_{X_2}(x_2)} \tag{11}$$

where $F(x_1, x_2) = P_r[X_1 \le x_1, X_2 \le x_2]$, $F_{X_1}(x_1) = P_r[X_1 \le x_1]$ and $F^*(x_1, x_2) = P_r[X_1 > x_1, X_2 > x_2]$.

## DATA

### Chukwooki and modern flip-bucket type rainfall data

The Chukwooki, an old rain gauge in Korea, was invented in the Choson Dynasty about 500 years ago. However, the longest rainfall data available are those collected in Seoul (see Figure 2) during the period 1777–1910. Jung (1999) has fully recovered the old Korean rainfall data collected by the Chukwooki, and translated them into the currently used scale units (hereafter it is called the Chukwooki data). In fact, the Chukwooki data were recorded with the Korean foot-rule such as Pun, Chi, and Cha, which approximately correspond to 2, 20, and 200 mm, respectively. Jung (1999) also proved that the Chukwooki data have enough accuracy, compared with modern measuring systems (Jung *et al.* 2001).

Jung (1999) explained that the Chukwooki data are a bit different from modern measurements in several aspects, such as the measuring units, measuring interval, and operation schedule. Also, the Chukwooki data do not contain the winter precipitation by snow (approximately 40 mm/year), and the rainfall below about 2 mm (approximately 35–40 mm/year). The sum of these two amounts becomes 75–80 mm/year. Another aspect of the Chukwooki data is the data collection time interval. Unlike the modern rainfall data, the data collection time interval was two hours in the Choson Dynasty; thus, one hour then is two hours in the modern clock system.

Modern flip-bucket type hourly rainfall data in Seoul have been collected since 1961 (hereafter it is called the modern data). Between 1911 and 1960, the rainfall data collected were mostly daily ones, along with sporadic 3-hour and 6-hour rainfalls. Due to the inconsistent structures of Chukwooki data, the modern data since 1961, and the data collected between 1911 and 1960, this study decided to analyze and compare the Chukwooki data and the modern hourly data collected since 1961. As most of the annual precipitation is concentrated on the wet summer season, and as the Chukwooki did not measure the winter


**Figure 2** | The location of the Chukwooki and modern rain gauge station in Seoul, Korea.

precipitation, only the wet season data (from June to September) were considered. Figure 3 shows the annual wet season rainfall data considered in this study. Here, the mean and standard deviation for the Chukwooki data are 852.1 mm and 360.96 mm, and those for the modern rain gauge data are 782.2 mm and 311.21 mm, respectively. Even though the Chukwooki could not measure the rainfall below about 2 mm, the mean value of the Chukwooki data was found to be slightly higher than the modern data.

As modern data were recorded hourly, it is not difficult to separate independent rainfall events using the given inter-event time definition (IETD). Separation of independent events is more straightforward, as the Chukwooki data were recorded as event-based, with its rainfall duration and total rainfall depth. However, as there are many cases where the inter-event period is less than the IETD, additional analysis should be done to derive the consistent independent rainfall events, as in the analysis of modern data. In Korea, there have been several studies to propose the proper IETD to secure the independent rainfall event, by following Restrepo-Posada & Eagleson (1982) and Adams & Papa (2000). In most studies, the IETD was decided to be about 10 hours (Lee & Jeong 1992; Kwon 2003).

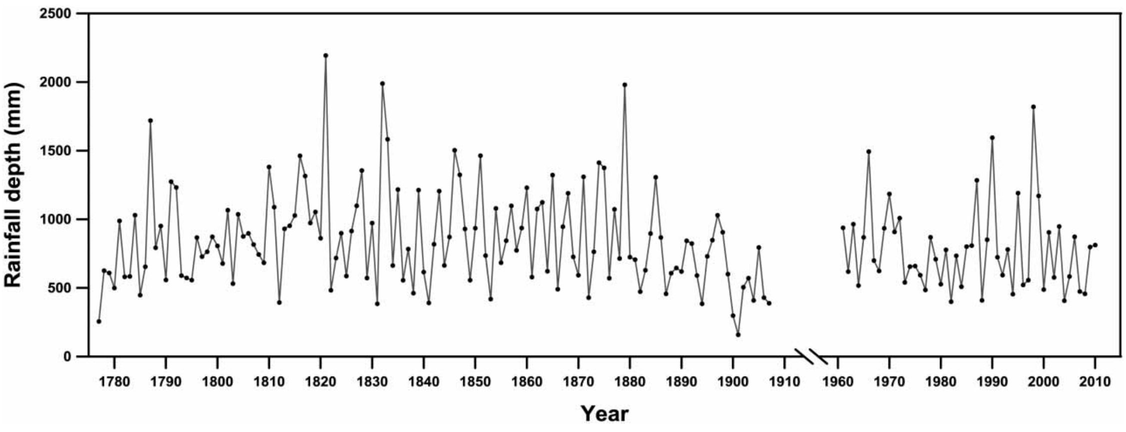


**Figure 3** | Annual wet season (from June to September) rainfall data considered in this study.

Another consideration to derive the independent rainfall events is the rainfall threshold. For the modern data, 0.5 mm is the generally applied rainfall threshold. However, as mentioned earlier, the Chukwooki could not measure the rainfall when its total depth was less than 2 mm. That is, for the modern data, independent rainfall events with total depth of less than 2 mm should be removed, to be fairly compared with the Chukwooki data. In this study, the independent rainfall events were derived by applying an IETD of 10 hours, and threshold of 2 mm. From now on, the rainfall events derived from the Chukwooki data and the modern data will be called the Chukwooki rainfall events and the modern rainfall events, respectively.

## Basic statistics of independent rainfall events

Table 1 summarizes and compares the annual mean numbers of independent rainfall events, along with other basic components of the independent rainfall events, such as the total rainfall depth, rainfall intensity and rainfall duration of the Chukwooki rainfall events and modern rainfall events. The correlation among components of the independent rainfall events were also derived and are compared separately in Table 2.

First, as given in Table 1, to see the effect of the rainfall threshold and data collection time interval, a total of four different sets of modern independent rainfall events were derived, and compared with the Chukwooki counterparts. These were made by applying two different rainfall thresholds (0 and 2 mm) and two different data collection time intervals (1 and 2 hours). As can be expected, the number of independent rainfall events decreases with the higher rainfall threshold. Increasing the rainfall threshold from 0 to 2 mm decreases the average number of rainfall events by more than 20%. On the other hand, the change of data collection time interval from 1 to 2 hours was found insignificant in this study. When applying both the rainfall threshold of 2 mm and the data collection time interval of 2 hours to the modern rainfall data, just as in the Chukwooki data, the average number of rainfall events decreased from 34.8 (the case when the rainfall threshold of 0 mm and the data collection time interval of 1 hour are applied) to 26.2.

The effect of the rainfall threshold and the data collection time interval was also found to be similar for each component of a rainfall event. For example, the total rainfall depth of a rainfall event was increased by applying the increased rainfall threshold of 2 mm. The increased data collection time interval was also found to increase the total rainfall depth a little. However, the rainfall duration has been decreased by half when applying the increased rainfall threshold of 2 mm, which also resulted in significant increase of the rainfall intensity by about three times. An increased data collection time interval also decreased the rainfall intensity, but no significant change could be found in the rainfall duration. In fact, this result is well expected, as minor independent rainfall events could be removed by applying the rainfall threshold of 2 mm. Also, the rainfall duration can be significantly increased, as most rainfall events are short-duration ones, of just one or two hours.

Among four cases compared, the rainfall events derived by applying the rainfall threshold of 2 mm and the data collection time interval of 2 hours were found to be most similar to the Chukwooki rainfall events. First of all, the annual mean numbers of independent rainfall events were found to be similar to each other, 25.2 for the Chukwooki data and 26.2 for the modern data. The total rainfall depths per rainfall event were also similar to each other,

**Table 1** | Basic statistics of independent rainfall events of the Chukwooki and modern rainfall data (standard deviation in brackets)

| | Chukwooki data | | | | Modern data | | | |
|---|---|---|---|---|---|---|---|---|
| Threshold (mm)/Data collection interval (h) | Annual mean number | Total depth (mm) | Mean intensity (mm/h) | Mean duration (h) | Annual mean number | Total depth (mm) | Mean intensity (mm/h) | Mean duration (h) |
| 0/1 | – | – | – | – | 34.8 (5.6) | 29.3 (53.3) | 1.8 (2.4) | 14.1 (16.1) |
| 0/2 | – | – | – | – | 34.3 (5.3) | 29.8 (53.7) | 1.6 (2.1) | 15.4 (16.4) |
| 2/1 | – | – | – | – | 25.6 (5.1) | 34.4 (50.4) | 5.3 (4.7) | 7.4 (9.0) |
| 2/2 | 25.2 (7.3) | 34.3 (45.0) | 4.9 (4.4) | 8.2 (8.8) | 26.2 (4.8) | 36.1 (52.8) | 3.8 (3.5) | 7.1 (8.2) |

**Table 2** | Correlation coefficients among components of independent rainfall events of the Chukwooki and modern rainfall data (for the same rainfall threshold 2 mm and data collection interval 2 hours)

| Components | Chukwooki data | Modern data |
|---|---|---|
| Total rainfall depth ~ Rainfall intensity | 0.17 | 0.44 |
| Total rainfall depth ~ Rainfall duration | 0.83 | 0.65 |
| Rainfall intensity ~ Rainfall duration | −0.15 | −0.03 |

34.3 for the Chukwooki rainfall events and 36.1 mm for the modern rainfall events. However, the Chukwooki rainfall events show somewhat higher rainfall intensity and longer rainfall duration than the modern rainfall events. From comparison of the standard deviations derived, it was also found that the variability of the Chukwooki rainfall events was much higher than that of the modern rainfall events.

The bivariate histograms for the total rainfall depth and the rainfall intensity of all the rainfall events, derived by applying the same rainfall threshold of 2 mm and the data collection interval of 2 hours, are given in Figure 4. As can be seen in this figure, the Chukwooki data and modern data are very similar in shape, and most of the independent rainfall events are concentrated in the region of small total rainfall depth and low rainfall intensity. Also, the frequency of the rainfall events decreases exponentially as the total rainfall depth or the rainfall intensity increases, in both cases of Chukwooki and modern data.

The correlations among those components are also useful to understand the possible change of the rainfall event characteristics. As can be seen in Table 2, the correlation between the total rainfall depth and the rainfall duration, which is very high in general, has loosened somewhat from 0.83 to 0.65. However, the correlation between the total rainfall depth and the rainfall intensity has become much stronger, from 0.17 to 0.44. The correlation between the rainfall intensity and rainfall duration has remained insignificant, but rather weakened, from −0.15 to −0.03. On the whole, the rather opposite behavior of the rainfall intensity and duration compared with the total rainfall depth of the Chukwooki and modern rainfall event data is noticeable.

As the correlation coefficients estimated among three components of rainfall events are all within the range of its application, it is possible to use any two components

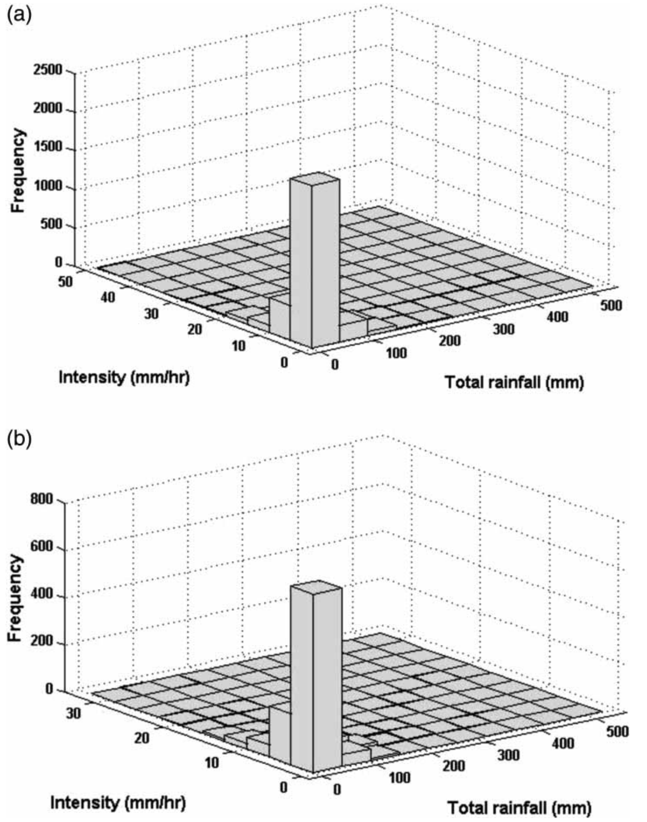


**Figure 4** | (a) Bivariate histograms of Chukwooki data. (b) Bivariate histograms of modern rainfall event data.

for the bivariate frequency analysis, using Freund's bivariate exponential distribution. In this study, the total rainfall depth and rainfall intensity were selected for the bivariate frequency analysis, to determine the annual maximum independent rainfall events. In fact, this decision was made as the correlation between the total rainfall depth and the rainfall intensity was positive, but not that high. It was also possible to use the total rainfall depth and the rainfall duration, but this was excluded as their correlation was too high. The negative correlation between the rainfall intensity and the rainfall duration was another reason to exclude this combination.

## COMPARISON OF ANNUAL MAXIMUM RAINFALL EVENT SERIES

The parameters of Freund's bivariate exponential distribution were estimated annually. The annually estimated parameters are known to better consider the different rainfall characteristics of wet and dry years. Park *et al.* (2014) also showed that the parameters estimated annually were better for considering the annual fluctuation of rainfall characteristics. As the number of independent rainfall events is more than 30 every year, there may be no serious problem in estimating the parameters annually.

The parameter estimation results are shown in Figure 5. As can be seen in this figure, the parameters estimated for the Chukwooki rainfall events and those for the modern rainfall events look very similar. Somewhat different patterns can be found around 1900, which was, in fact, an exceptional period, with abnormally low annual rainfall amount in far eastern Asia, including China and Japan (Jung 1999; Jung *et al.* 2001). During this period, the annual rainfall amount was less than two-thirds that of a normal year.

It is also worthwhile to review the relation between the characteristics of selected annual maximum rainfall events, and the estimated parameters of Freund's bivariate exponential distribution. In particular, it was found that the parameters related to the total rainfall depth become higher when the annual rainfall depth (in fact, during the wet summer season) is high. On the other hand, the parameters related to the rainfall intensity become higher when the annual rainfall depth is low. The correlation coefficient between the total rainfall depth and the rainfall intensity was found insignificant in the parameter

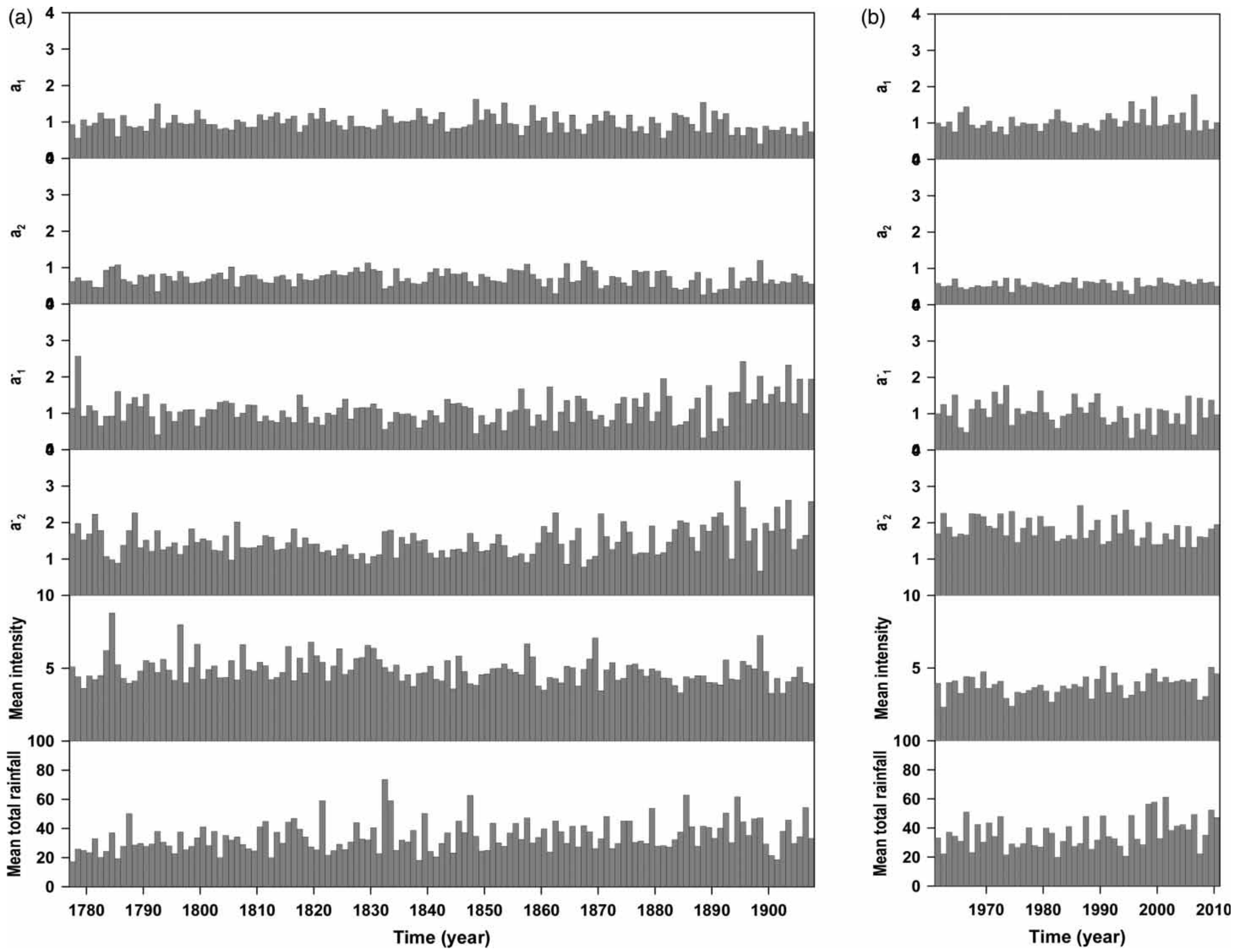


**Figure 5** | Estimated parameters of the Freund's bivariate exponential distribution and their comparison with the mean rainfall intensity and total rainfall depth of (a) Chukwooki data and (b) modern rainfall event data.

estimation. This result indicates that the impact of the total rainfall depth of a rainfall event becomes more significant during wet years in deciding the annual maximum rainfall event, but the rainfall intensity becomes more significant during dry years. Quite interestingly, this result is also coincident with the relative amount of direct runoff or flood during the wet and dry years. For example, during a dry year, the rainfall intensity must be much higher than the infiltration rate to cause enough direct runoff to cause a flood. On the other hand, as the soil must be wet during a wet year, the total rainfall depth, rather than the rainfall intensity, is more closely related to a flood.

The annual maximum rainfall event was decided to be the one that gives the longest return period. Figure 6 shows the components of the derived annual maximum rainfall event series. It is quite interesting that the temporal variability of each component of the annual maximum rainfall event series is very high compared with those of the estimated parameters, as shown in Figure 5. This is also the same in both the Chukwooki and modern rainfall event series. This finding is confirmed in Table 3, with the estimated standard deviation.

Table 3 summarizes the basic statistics of the annual maximum rainfall event. Mean values and standard deviations of total rainfall depths, rainfall intensities, and rainfall durations of the independent rainfall event series constructed are summarized in this table. As can be seen in Table 3, the total rainfall depth per annual maximum rainfall event has increased significantly, from 144.9 mm for the Chukwooki data to 172.1 mm for the modern data. This result is very interesting, as the rainfall intensity has decreased a little, but the rainfall duration has remained the same, without any significant change. In fact, the answer is very straightforward, because the mean of the total rainfall depth data can be different from the multiplication of the mean rainfall intensity and the mean rainfall duration. The difference can be more significant if the distributions of the rainfall intensity and the rainfall duration are highly skewed.

It is also interesting to compare the statistics of all the rainfall events and the annual maximum rainfall events.

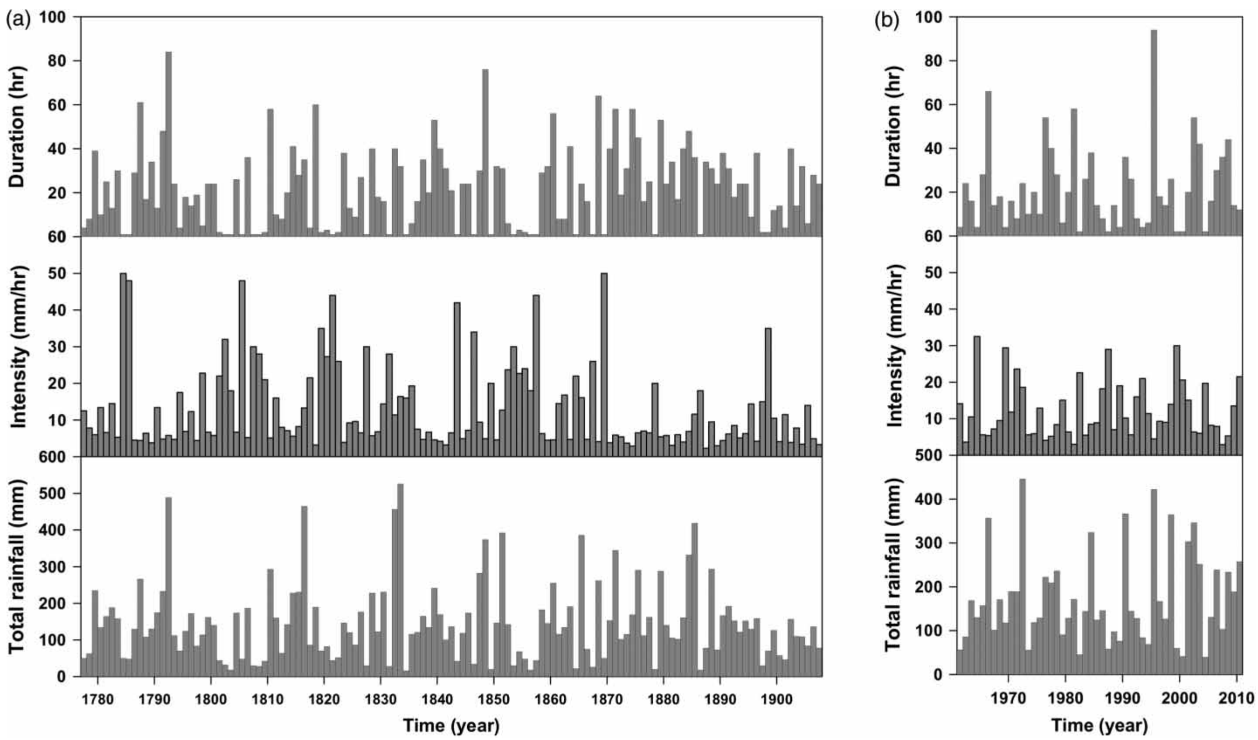


**Figure 6** | Time series plot of components of annual maximum rainfall event series derived from (a) Chukwooki data and (b) modern rainfall data.

**Table 3** | Basic statistics of annual maximum independent rainfall events of the Chukwooki and modern rainfall data for the rainfall threshold 2 mm and the data collection time interval 2 hours (standard deviation in brackets)

| | Chukwooki data | Modern data |
|---|---|---|
| Total rainfall depth (mm) | 144.9 (106.0) | 172.1 (104.0) |
| Mean intensity (mm/h) | 13.1 (11.5) | 12.3 (7.8) |
| Mean duration (h) | 21.8 (18.4) | 21.7 (19.2) |

For the Chukwooki data, both the rainfall intensity and rainfall duration of the annual maximum rainfall events were estimated to be about 2.7 times higher than those of all the events, and the total rainfall depth was greater by about 4.2 times. The trend of the modern data was also similar to the Chukwooki data. The rainfall intensity and the rainfall duration of the annual maximum rainfall events were estimated to be about 3.2 and 2.6 times higher than those of all the rainfall event data, respectively. The total rainfall depth of the annual maximum rainfall events was greater by about 4.8 times than that of all the rainfall events.

The correlations among three components of the annual maximum independent rainfall events were estimated to be very different from those of all the independent rainfall events (Table 4). The correlation between the total rainfall depth and the rainfall duration was found to be very strong, as in all the rainfall events, at 0.75 for the Chukwooki data and 0.70 for the modern data. Also, the correlation between the rainfall intensity and the rainfall duration was found to be very strong, with −0.67 for the Chukwooki data and −0.65 for the modern data. Keeping in mind that this correlation was insignificant in the case of all rainfall events, this particular relation is more noticeable. The negative correlation between the total rainfall depth and the rainfall intensity, which is −0.41 for the Chukwooki data and −0.27 for the modern data, is also noticeable. This correlation was positive in the case of all rainfall events but not for the annual maximum rainfall events.

**Table 4** | Correlation coefficients among components of annual maximum independent rainfall events of the Chukwooki and modern rainfall data (for the same rainfall threshold 2 mm and data collection interval 2 hours)

| Components | Chukwooki data | Modern data |
|---|---|---|
| Total rainfall depth ~ Rainfall intensity | −0.41 | −0.27 |
| Total rainfall depth ~ Rainfall duration | 0.75 | 0.70 |
| Rainfall intensity ~ Rainfall duration | −0.67 | −0.65 |

This rather opposite result in the derived correlations is, in fact, mainly due to the two variables selected for the bivariate frequency analysis. In fact, this result explains what kind of rainfall events were selected as the annual maximum ones. As explained earlier, the only criterion to determine the annual maximum rainfall event is the return period. Also, bivariate frequency analysis was done for the variables of total rainfall depth and rainfall intensity. Thus, to have the longest return period, a short-duration rainfall event, as it may not have a large total rainfall depth, should have very high rainfall intensity. On the other hand, a long-duration rainfall event could have a large total rainfall depth, even though the rainfall intensity is not so high. Thus, the correlation between the rainfall intensity and rainfall duration becomes more strongly and negatively proportional. Also, the correlation between the total rainfall depth and rainfall intensity could not be positive.

It is also interesting to visually compare the annual maximum rainfall event series by representing them with rectangular pulses (Figure 7). It is also helpful to compare the annual maximum Chukwooki rainfall events, along with the modern ones. First of all, it was found that various rectangular pulses were selected as the annual maximum. Some of them are those with very high rainfall intensity, and others are those with very long rainfall duration. However, their overall trend, from high rainfall intensity to long rainfall duration, is not fully exponential (in its decay pattern). There can be found many rectangular pulses with both high rainfall intensity and long rainfall duration. In fact, these exceptional rainfall events are those with rather long return periods compared with the others. These characteristics are the same in both the Chukwooki data and the modern data. However, it is not easy to distinguish one from the other in Figure 7.

## BIVARIATE FREQUENCY ANALYSIS OF THE ANNUAL MAXIMUM RAINFALL EVENT SERIES

The bivariate histogram of the annual maximum rainfall event series is given in Figure 8. This histogram does

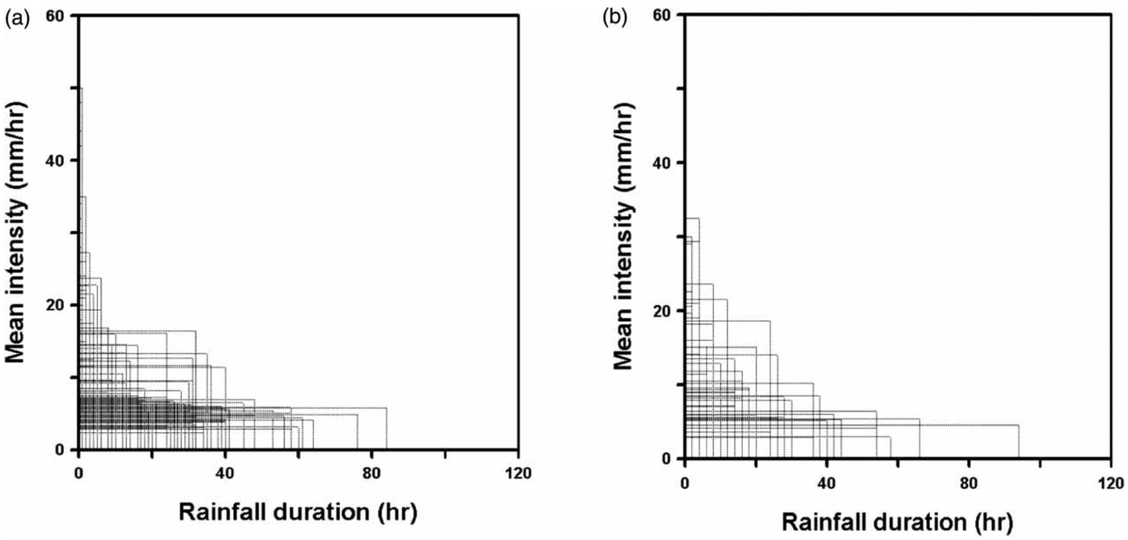


**Figure 7** | The annual maximum rainfall event series overlapped from (a) Chukwooki data and (b) modern rainfall data.

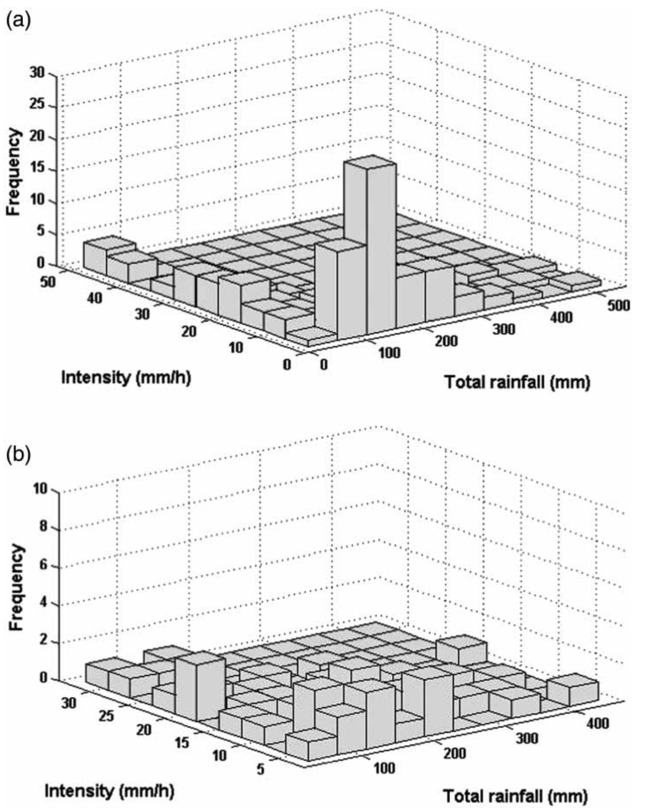


**Figure 8** | Bivariate histograms of the annual maximum rainfall events of (a) Chukwooki data and (b) modern rainfall data.

not show any exponential decay pattern, but is, in fact, better fitted to the bivariate extreme distribution (Park *et al.* 2014).

The parameters of the bivariate logistic model were estimated for the annual maximum rainfall event series constructed in the previous section. The maximum likelihood method was used for the parameter estimation, as introduced in the previous section. The estimates of the location, scale and shape parameters for a marginal variable, the total rainfall depth, were estimated to be 96.7, 77.6 and −0.042 for the Chukwooki data, and 126.3, 82.8, and 0.024 for the modern rain gauge data, respectively. Also, the estimates of the location, scale and shape parameters for a marginal variable, the rainfall intensity, were estimated to be 7.8, 8.1 and −0.068 for the Chukwooki data, and 8.9, 6.2, and 0.023 for the modern rain gauge data, respectively. The association parameters were estimated to be about 0.90 for the Chukwooki data, and 0.81 for the modern rain gauge data.

The empirical marginal cumulative probabilities of the total rainfall depth and the rainfall intensity were compared with the theoretical cumulative probabilities. Figure 9 shows the results for the Chukwooki and modern rain gauge data. Here, the dotted lines indicate the upper and lower limits, based on the critical values of the Kolmogorov-Smirnov statistics at a significance level of 5%, which shows that the bivariate logistic model was very acceptable. In particular, the observed and theoretical probabilities for all cases were found to agree well at higher tails, which represent the extreme rainfall events. However, the rainfall intensity

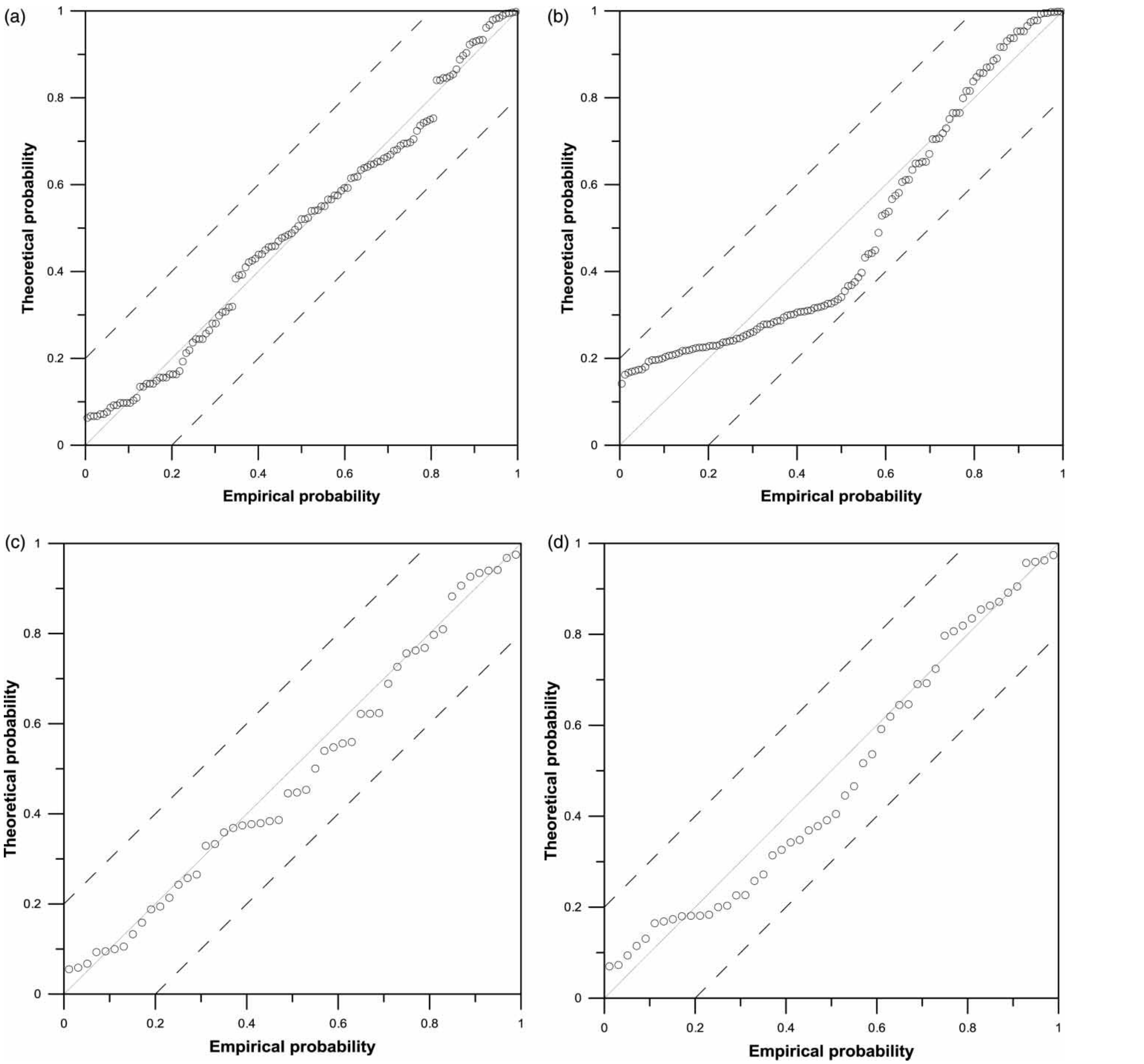


**Figure 9** | Empirical marginal cumulative probabilities of (a) the total rainfall depth and (b) the rainfall intensity of Chukwooki events; and (c) the total rainfall depth and (d) the rainfall intensity of modern rainfall events (the dotted lines indicate the upper and lower limits based on the critical values of the Kolmogorov-Smirnov statistics at a significance level of 5%).

data of the Chukwooki rainfall events shows some disagreement with the theoretical ones in the lower tails as shown in Figure 9(b).

For given return periods, rainfall events were estimated, based on the concept of AND joint return period, and compared for the Chukwooki data and modern rain gauge data. The AND concept was adopted because it was found to be more consistent with conventional univariate frequency analysis (Park et al. 2014). The results are summarized in Figure 10 (Table 5 also compares the results for the rainfall durations of 6, 12, 24, 48 and 72 hours). As can be seen in Figure 10, both results are similar for the rainfall durations shorter than 24 hours. However, from the rainfall duration of 24 hours, the total rainfall depth per rainfall event of

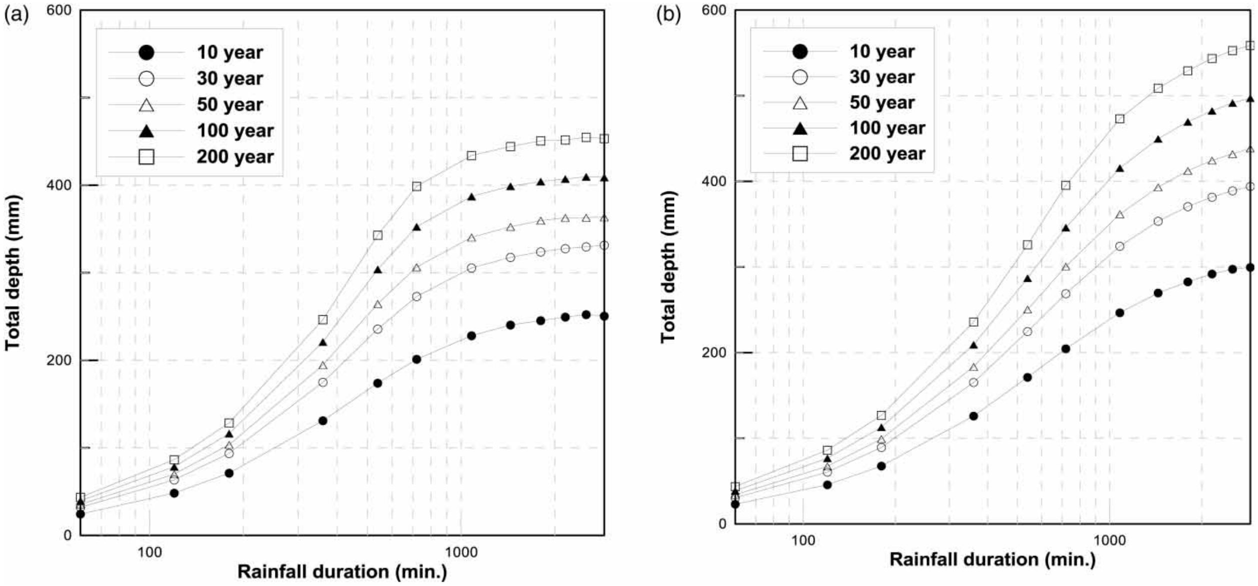


**Figure 10** | Total rainfall depth per rainfall event for given return period and rainfall duration of (a) Chukwooki data and (b) modern rainfall data.

the modern data becomes higher than that of the Chukwooki data. The difference becomes larger depending on the rainfall duration, from about 10% for the rainfall duration of 24 hours to about 25% for the rainfall duration of 72 hours. For example, for the rainfall duration of 24 hours, the applications of the Chukwooki data and the modern rain gauge data resulted in total rainfall depths of 240.3 and 269.7 mm (rainfall intensities of 10.0 mm/h and 11.2 mm/h) for the return period of 10 years, respectively. The difference between the two was about 10%. As another case for the rainfall duration of 72 hours, the applications of the Chukwooki data and the modern rain gauge data resulted in total rainfall depths of 410.4 mm and 510.6 mm (rainfall intensities of 5.7 mm/h and 7.1 mm/h) for the return period of 100 years, respectively. The difference was about 20% in this case.

Using the graphs in Figure 10, the return period of a rainfall event can be assigned. The heaviest rainfall event that occurred during the Chukwooki period was the one in 1885. Its duration was 96 hours, and the total rainfall depth was 565.9 mm. The return period estimated was about 500 years. The second heaviest rainfall event occurred in 1832, whose duration was 88 hours, and the total rainfall depth was 516.0 mm. The return period of this rainfall event

**Table 5** | Comparison of total rainfall depths (mm) derived by the bivariate frequency analysis of the Chukwooki (Chuk.) and modern rain gauge (Mod.) data for the rainfall durations of 6, 12, 24, 48 and 72 hours

| | Rainfall duration (h) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 6 | | 12 | | 24 | | 48 | | 72 | |
| Return period (year) | Chuk. | Mod. | Chuk. | Mod. | Chuk. | Mod. | Chuk. | Mod. | Chuk. | Mod. |
| 2 | 54.0 | 61.3 | 83.2 | 101.4 | 102.4 | 133.3 | 108.6 | 147.3 | 109.6 | 150.7 |
| 10 | 131.0 | 126.0 | 201.2 | 204.6 | 240.3 | 269.7 | 250.5 | 299.6 | 251.7 | 307.2 |
| 30 | 174.9 | 165.2 | 272.9 | 268.9 | 317.6 | 353.5 | 331.5 | 394.1 | 334.4 | 405.6 |
| 50 | 194.7 | 183.6 | 306.5 | 300.7 | 352.5 | 393.2 | 363.4 | 438.6 | 364.9 | 452.4 |
| 100 | 221.0 | 209.3 | 352.6 | 346.1 | 398.9 | 449.7 | 409.0 | 497.3 | 410.4 | 510.6 |
| 200 | 246.5 | 235.8 | 398.7 | 395.4 | 444.2 | 508.9 | 453.3 | 558.8 | 454.5 | 572.4 |

was estimated to be about 400 years. The third heaviest rainfall event occurred in 1865, its duration was 24 hours, and the total rainfall depth was 385.4 mm. The return period was estimated to be about 100 years. Similar severe rainfall events have also been recorded in the modern rain gauge period. The heaviest rainfall event was recorded in 1998. Its duration was 77 hours and the total rainfall depth was 580.5 mm. The return period estimated was about 400 years. The second rainfall event was occurred in 1972; its duration was 47 hours and total rainfall depth was 455.8 mm. The return period was estimated to be about 200 years. The third one occurred in 1990, with a duration of 36 hours and total rainfall depth of 367.5 mm. The return period was estimated to be about 100 years. These three rainfall events were recorded as catastrophic ones, resulting in severe casualties and serious property losses (Seoul Metropolitan City 1999).

Simply comparing the return period of those severe events in the Chukwooki period and the modern rain gauge period, it was found that there is no significant difference in the size of the rainfall events. Considering the difference in the data length of the Chukwooki data and the modern data, the difference in the return periods of the heaviest rainfall events could be easily understandable.

## CONCLUSIONS

In this study, the annual maximum rainfall event series were constructed and compared, for both the modern flip-bucket type rainfall data collected since 1961 (the modern data) and the old Chukwooki rainfall data collected from 1777 to 1910 (the Chukwooki data). First, independent rainfall events were derived by applying the same rainfall threshold of 2 mm and data collection time interval of 2 hours to both the Chukwooki and the modern data. Annual maximum rainfall event series were then constructed by applying Freund's bivariate exponential distribution annually. Finally, bivariate frequency analysis was done for the annual maximum rainfall event series, constructed by applying the bivariate logistic model to evaluate and quantify their characteristics. The results can be summarized as follows:

First, the annual mean numbers of independent rainfall events were found to be similar to each other: 25.2 for the Chukwooki data and 26.2 for the modern data. The total rainfall depths per rainfall event were also similar to each other: 34.3 for the Chukwooki rainfall events and 36.1 mm for the modern rainfall events, respectively. However, the Chukwooki rainfall events show a somewhat higher rainfall intensity and longer rainfall duration than the modern rainfall events. The correlation between the total rainfall depth and the rainfall duration has loosened a bit, from 0.83 for the Chukwooki data to 0.65 for the modern data. However, the correlation between the total rainfall depth and rainfall intensity has become much stronger, from 0.17 to 0.44. The correlation between the rainfall intensity and rainfall duration has remained insignificant, but rather weakened, from $-0.15$ to $-0.03$.

Second, the total rainfall depth per annual maximum rainfall event has increased rather significantly, from 144.9 mm for the Chukwooki data to 172.1 mm for the modern data. This result is very interesting, as the rainfall intensity has decreased a little, but the rainfall duration has remained the same, without any significant change. The correlation between the total rainfall depth and rainfall duration was found to be very strong, as in all the rainfall events, at 0.75 for the Chukwooki data and 0.70 for the modern data. Also, the correlation between the rainfall intensity and the rainfall duration was found to be very strong, at $-0.61$ for the Chukwooki data and $-0.65$ for the modern data. The negative correlation between the total rainfall depth and rainfall intensity, which is $-0.41$ for the Chukwooki data and $-0.27$ for the modern data, is also noticeable.

Third, bivariate frequency analysis of the annual maximum rainfall event series shows that the total rainfall depth per rainfall event is similar for rainfall durations shorter than 24 hours. However, from the rainfall duration of 24 hours, the total rainfall depth per rainfall event of the modern data becomes higher than that of the Chukwooki data. The difference becomes larger depending on the rainfall duration, from about 10% for the rainfall duration of 24 hours to about 25% for the rainfall duration of 72 hours. The heaviest rainfall event occurring during the Chukwooki rainfall period was the one in 1885. With its duration of 96 hours and total rainfall depth of 565.9 mm, the return period was estimated to be about 500 years. Also, the heaviest rainfall event in the modern rainfall period occurred in 1998. The duration of this event was 77 hours and the total rainfall depth was 580.5 mm. The return period was estimated to be about 400 years.

As can be found in the above summary of the results derived by comparing the Chukwooki rainfall events and the modern rainfall events, they are not so different. The difference can even be insignificant if considering the difference between two measuring devices, the Chukwooki and modern rain gauge. It should also be noticed that this study was based on the 2-hour data, thus the results derived in this study cannot explain the possible change of very-short-duration rainfall events, especially those of less than two hours. This must be a limitation of this study. However, despite the fact that the characteristics of rainfall events have changed slightly, it may not be justified to conclude that the rainfall characteristics have changed significantly during the last 200 years.

## ACKNOWLEDGEMENTS

This study was initiated by the support from Basic Science Research Program through the Korea Research Foundation (KRF-2008-313-D01083) and National Research Foundation of Korea (NRF) through the Ministry of Education, Science and Technology (No. 2010-0014566).

## REFERENCES

Adams, B. J. & Papa, F. 2000 *Urban Storm Water Management Planning with Analytical Probabilistic Models.* John Wiley & Sons, New York, USA.

Bacchi, B., Becciu, G. & Kottegoga, N. T. 1994 Bivariate exponential model applied to intensities and durations of extreme rainfall. *Journal of Hydrology* **155**, 225–236.

Coles, S. 1993 Regional modeling of extreme storms via max-stable processes. *Journal of the Royal Statistical Society* **55** (4), 797–816.

Coles, S. & Tawn, J. A. 1991 Modeling extreme multivariate events. *Journal of the Royal Statistical Society: Series B* **53** (2), 377–392.

De Michele, C., Salvadori, G., Canossi, M., Petaccia, A. & Rosso, R. 2005 Bivariate statistical approach to check adequacy of dam spillway. *Journal of Hydrologic Engineering* **10** (1), 50–57.

Frei, C. & Schăr, C. 2001 Detection probability of trends in rare events: theory and application to heavy precipitation in the Alpine region. *Journal of Climatology* **14**, 1568–1584.

Freund, J. 1961 A bivariate extension of the exponential distribution. *Journal of American Statistical Association* **56**, 971–977.

Gergis, J. & Fowler, A. 2005 Classification of synchronous oceanic and atmospheric El Niňo-Southern Oscillation (ENSO) events for palaeoclimatic reconstruction. *International Journal of Climatology* **25**, 1541–1565.

Gumbel, E. J. 1960 Bivariate exponential distributions. *Journal of American Statistical Association* **55**, 698–707.

Gumbel, E. J. 1965 Two systems of bivariate extremal distribution. In: *35th Session of the International Statistical Institute, Beograd, No. 69*. International Statistical Institute.

Jhun, J. G. & Moon, B. K. 1997 Restorations and analyses of rainfall amount observed by Chukwooki. *Asia-Pacific Journal of Atmospheric Sciences* **33** (4), 691–707.

Jung, H. S. 1999 *Interpretation of the Transient Variation in the Time Series of Precipitation Amounts in Seoul*. PhD Thesis, Seoul National University, Seoul, Korea.

Jung, H. S., Lim, G. H. & Oh, J. H. 2001 Interpretation of the transient variations in the time series of precipitation amounts in Seoul, Korea: Part 1. Diurnal variation. *Journal of Climate* **14** (13), 2989–3004.

Justel, A., Peña, D. & Zamar, R. 1997 A multivariate Kolmogorov-Smirnov test of goodness of fit. *Statistics & Probability Letters* **35**, 251–259.

Kim, D., Yoo, C. & Kim, H. J. 2007 Evaluation of major storm events both measured by Chukwooki and recorded in annals of Chosen Dynasty: 2. Quantitative approach. *Journal of Korea Water Resources Association* **40** (7), 545–554.

KMA 2004 *Modern Meteorology 100 Years History.* Korea Meteorological Administration, Seoul, Korea.

Kotz, S., Balakrishnan, N. & Johnson, N. L. 2000 *Continuous Multivariate Distributions Volume 1: Models and Applications.* John Wiley & Sons, New York.

Kwon, J. H. 2003 *Rainfall Analysis to Estimate the Amount of Nonpoint Source Pollution*. MS Thesis, Korea University, Seoul, Korea.

Lee, D. R. & Jeong, S. M. 1992 Spatial-temporal characteristics of rainfall in the Han river basins. *Journal of Korean Association of Hydrology Science* **25** (4), 75–85.

Lee, C. H., Kim, T. W., Chung, G., Choi, M. & Yoo, C. 2010 Application of bivariate frequency analysis to the derivation of rainfall-frequency curves. *Stochastic Environmental Research and Risk Assessment* **24**, 389–397.

Muller, A., Bacro, J. N. & Lang, M. 2008 Bayesian comparison of different rainfall depth-duration-frequency relationships. *Stochastic Environmental Research and Risk Assessment* **22**, 33–46.

Park, M., Yoo, C., Kim, H. & Jun, C. 2014 Bivariate frequency analysis of annual maximum storm event series in Seoul, Korea. *Journal of Hydrologic Engineering* **19** (6), 1080–1088.

Rao, A. R. & Hamed, K. H. 2000 *Flood Frequency Analysis.* CRC Press, Boca Raton, FL.

Restrepo-Posada, P. J. & Eagleson, P. S. 1982 Identification of independent rainstorms. *Journal of Hydrology* **55**, 303–319.

Salvadori, G. & De Michele, C. 2007 On the use of copulas in hydrology: theory and practice. *Journal of Hydrologic Engineering* **12** (4), 369–380.

Seoul Metropolitan City 1999 *Flood Disasters in Seoul*. Korea Water Resources Association, Seoul, Korea.

Serinaldi, F. & Grimaldi, S. 2007 Fully nested 3-copula: procedure and application on hydrological data. *Journal of Hydrologic Engineering* **12** (4), 420–430.

Shiau, J. T. 2003 Return period of bivariate distributed extreme hydrological event. *Stochastic Environmental Research and Risk Assessment* **17**, 42–57.

Smith, E. 2005 *Bayesian Modeling of Extreme Rainfall Data*. PhD Thesis, University of Newcastle upon Tyne, UK.

Stedinger, J. R. & Cohn, T. A. 1986 Flood frequency analysis with historical and palaeoflood information. *Water Resources Research* **22** (5), 785–793.

Tawn, J. A. 1988 Bivariate extreme value theory: models and estimation. *Biometrica* **75** (3), 397–415.

Tawn, J. A. 1990 Modeling multivariate extreme value distribution. *Biometrica* **77** (2), 245–253.

Vannitsem, S. & Naveau, P. 2007 Spatial dependences among precipitation maxima over Belgium. *Nonlinear Processes in Geophysics* **14**, 621–630.

Wang, B., Ding, Q. & Jhun, J. G. 2006 Trends in Seoul (1778–2004) summer precipitation. *Geophysical Research Letters* **33**, L15803.

Yoo, C. 2006 Long-term analysis of wet and dry years in Seoul, Korea. *Journal of Hydrology* **318** (1–4), 24–36.

Yoo, C. 2008 *The 5th Year Report: Rainfall Analysis and Prediction Technology in Urban Areas (Estimation of Changes in Rainfall Characteristics due to Climate Change and Urbanization)*. Urban Flood Management Research Center, Seoul, Korea.

Yoo, C., Kim, D. & Kim, H. J. 2007 Evaluation of major storm events both measured by Chukwooki and recorded in annals of Chosen Dynasty: 1. Qualitative approach. *J. Korea Water Resources Association* **40** (7), 533–543.

Yue, S. 2001a The Gumbel logistic model for representing a multivariate storm event. *Advances in Water Resources* **24** (2), 179–185.

Yue, S. 2001b A bivariate gamma distribution for use in multivariate flood frequency analysis. *Hydrological Processes* **15** (6), 1033–1045.

Yue, S. 2001c A statistical measure of severity of El Niňo events. *Stochastic Environmental Research and Risk Assessment* **15**, 153–172.

Yue, S. & Rasmussen, P. 2002 Bivariate frequency analysis: discussion of some useful concepts in hydrological application. *Hydrological Processes* **16**, 2881–2898.

Yue, S. & Wang, C. Y. 2004 A comparison of two bivariate extreme value distributions. *Stochastic Environmental Research and Risk Assessment* **18** (2), 61–66.

Yue, S., Ouarda, T. B. M. J., Bobée, B., Gegendre, P. & Bruneau, P. 1999 The Gumbel mixed model. *Journal of Hydrology* **226** (1–2), 88–100.

First received 3 July 2016; accepted in revised form 24 August 2017. Available online 6 October 2017