METEOROLOGICAL APPLICATIONS
Meteorol. Appl. **24**: 31–42 (2017)
Published online 6 January 2017 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/met.1602

RMetS
Royal Meteorological Society

Consideration of documentary records in the Annals of the Choson Dynasty for the frequency analysis of rainfall in Seoul, Korea

Minkyu Park,[a] Chulsang Yoo[b]* and Changhyun Jun[c]

[a] *Department of Disaster Mitigation and Safety Science, Faculty of Convergence Science, Jungwon University, Chungcheongbuk-do, Korea*
[b] *Department of Civil, Environmental and Architectural Engineering, College of Engineering, Korea University, Seoul, Korea*
[c] *Department of Civil and Environmental Engineering, School of Engineering, Hong Kong University of Science and Technology, Kowloon, Hong Kong*

ABSTRACT: In this study, flood records in the Annals of the Choson Dynasty (1392–1910) were considered for the frequency analysis of rainfall in Seoul, Korea. The old Korean rain gauge (Chukwooki) data (1777–1907) were also used to evaluate major flood records in the Annals of the Choson Dynasty. The design rainfalls were estimated using the binomial censored data maximum likelihood estimator. The results can be summarized as follows. (1) The frequency analysis was done for 24 h rainfall amounts as this was assumed to represent the three different rainfall records well: the documentary records in the Annals of the Choson Dynasty, the Chukwooki data and the modern data. Also, given the possible uncertainty level of the classification results of the documentary records, only 19 catastrophic events (the highest level of flood) were considered in the frequency analysis. The threshold value was decided to be 388 mm by comparing the documentary records with the Chukwooki data. (2) The sensitivity analysis shows that the effects of the threshold value, the number of catastrophic records, and the accuracy level of the flood records in the Annals of the Choson Dynasty on the frequency analysis results are very limited. That is, the length of the flood records in the Annals of the Choson Dynasty was found to be long enough to overcome the possible uncertainty of the flood records. (3) The design rainfalls estimated by taking into account the flood records in the Annals of the Choson Dynasty were found to be very similar to those for the modern data. However, the confidence intervals estimated by considering all the records available were found to be just one third of those of the modern data.

KEY WORDS Annals of the Choson Dynasty; Chukwooki; rainfall data; frequency analysis; binomial censored data MLE; design rainfalls

Received 3 July 2015; Revised 24 June 2016; Accepted 24 June 2016

1. Introduction

When estimating a design rainfall or flood, especially if the return period is very much longer than the record length, a question may be raised about its confidence level (Stedinger and Lu, 1995; Reed, 1999). Korea, too, has this problem of insufficient data for the frequency analysis of rainfall or runoff. While the return period applied to important hydraulic structures is generally longer than 100 years, the data length available is less than 50 years in most locations. In this case, historical data such as record documents and flood traces can lessen the problem by improving the accuracy of the frequency analysis (Tasker and Thomas, 1978; Stedinger and Cohn, 1986; Frances, 1998).

With regard to this problem, much interest has been given to palaeohydrology worldwide. For example, the Working Group 4 of Japan, in the programme of Prediction in Ungauged Basins (PUB), International Association of Hydrological Sciences (IAHS), showed research on the estimation of extreme events considering historical records (Sivapalan *et al.*, 2003). In the United States, a database on hurricanes in historical records has been built and used at the National Hurricane Center. The European Union also launched a project called ‘Palaeofloods, Historical Data and Climatic Variability: Applications in Flood Risk Assessment’ to collect and use palaeoflood data for the analysis of current and future potential flood damage. The project ‘Systematic Palaeoflood and Historical Data for the Improvement of Flood Risk Estimation’ is another example in the European Union. China also keeps detailed information about typhoons in historical records. Liu *et al.* (2001) reported the frequency of typhoons in Guangdong, China, for 935 years (from 975 to 1909) by analysing related records in a local newspaper of Fang Zhi, China. All the historical weather data available in the Chinese history, including those from the inscriptions on bones and tortoise carapaces, were collected and organized by the China Meteorological Administration.

In Korea, there have been several studies on palaeohydrology, one of which is related to the old Korean rain gauge (Chukwooki) data (from 1777 to 1907) and another to the documentary records in the Annals of the Choson Dynasty (from 1392 to 1910). Originally, most studies of the Chukwooki data have focused on their validation (Jung, 1999) and comparison with modern flip-bucket type rain gauge data (Yoo, 2006; Yoo *et al.*, 2015a). However, recent studies have concentrated on the climate in the Chukwooki data period (Jung *et al.*, 2001; Wang *et al.*, 2006). Nowadays, researchers try to use the Chukwooki data for more practical purposes such as frequency analysis (Kim *et al.*, 2007a; Park *et al.*, 2014) and climate change (Park, 2009; Hwang *et al.*, 2010). On the other hand, studies related to the documentary records in the Annals of the Choson Dynasty have been limited to

* Correspondence: C. Yoo, Department of Civil, Environmental and Architectural Engineering, College of Engineering, Korea University, Seoul, 02841, Korea. E-mail: envchul@korea.ac.kr

© 2017 Royal Meteorological Society

their classification only. The classification of the rainfall events in the Annals of the Choson Dynasty has just been completed (Kim, 2000; Yoo *et al.*, 2015a) and will be used for many other related studies. In fact, this study is one of the possible applications of the classification result, and the authors are interested in the frequency analysis of the rainfall data including the classification results of the documentary records in the Annals of the Choson Dynasty.

The documentary records in the Annals of the Choson Dynasty are considered to be big rainfall events, as they were chosen by historiographers. Thus, these documentary records can be censored data in statistical terms exceeding a specific threshold value (Stedinger and Cohn, 1986). There have been many studies on how to factor these historical data into conventional univariate frequency analysis. A method for frequency analysis using historical data is the moment adjustment method proposed by the United States Water Resources Council Hydrology Committee (USWRC) (1967, 1982). This method is based on the concept of a threshold value, and a weight is given to each datum to obtain a corrected moment estimate (Kirby, 1981). A second method is based on the maximum likelihood estimator (MLE) for censored data. Developed by Hald (1949) and Cohen (1950, 1976), it has been applied to the parameter estimation of Gumbel, log-normal and log-Pearson type III distributions (Leese, 1973; Condie and Lee, 1982; Condie and Pilon, 1983). Recently, Stedinger and Cohn (1986) proposed an MLE for binomial censored data or the binomial censored data MLE to enable the use of qualitative data (i.e. the qualitative information in historical data) for frequency analysis.

In the present study, the flood records in the Annals of the Choson Dynasty (1392–1910) are reviewed and considered for the frequency analysis of rainfall in Seoul, Korea. The old Korean rain gauge (Chukwooki) data (1777–1907) were also used to derive the threshold for considering important flood records in the Annals of the Choson Dynasty. The total length of the data used becomes 619 years, dating from 1392 to 2010. Finally, design rainfalls are estimated using the binomial censored data MLE. Additionally, the sensitivity of the design rainfalls to the threshold value, the number of flood records considered and the accuracy level of the flood records is analysed to evaluate the quality of the frequency analysis.

This paper is divided into five sections, including the introduction and conclusions. The following section deals with the theoretical background on the binomial censored data MLE. The third section covers the characteristics of rainfall events recorded in the Annals of the Choson Dynasty and those recorded by the Chukwooki. The frequency analysis and sensitivity analysis are covered in the Section 4.

2. Binomial censored data MLE

Commonly used methods for frequency analysis using historical data include the moment adjustment (USWRC, 1967, 1982), the censored data MLE and the binomial censored data MLE (Bishop *et al.*, 1975; Stedinger and Cohn, 1986), all of which enable the re-estimation of model parameters used for the frequency analysis. Among these methods, the methods based on the MLE concept were found to be superior to the moment adjustment method (Stedinger and Cohn, 1986). In particular, better results could be obtained by applying the MLE concept when a higher threshold value was used. It was found that, as the threshold value becomes higher, the actual value becomes less important. Kim *et al.* (2007b) also compared the censored data MLE and the binomial censored data MLE by applying them to the rain gauge data in Korea. They concluded that the probable rainfall estimated by the censored data MLE is not so different from that estimated by the binomial censored data MLE and they also found that the estimated design rainfalls became more similar to each other as the threshold became higher. These results indicate that highly reliable probable rainfall can be obtained by applying either the censored data MLE or the binomial censored data MLE.

In the present study, the binomial censored data MLE were selected as a method of frequency analysis due to the qualitative nature of the historical data considered. Also, the log-normal distribution was considered in the frequency analysis because more theoretical research results are available than any others. It is also true that the log-normal distribution has long been used for the frequency analysis of rainfall (Condie and Lee, 1982; Kuczera, 1999; Reis and Stedinger, 2005).

The binomial censored data MLE of the log-normal distribution can be summarized as follows. Assume that s modern observations $\{X|X_1, X_2, \cdots, X_s\}$ and k historical observations $\{Y|Y_1, Y_2, \cdots, Y_k\}$ are available. The historical observations are all over the threshold value X_0 during the historical record length h. The likelihood function for the location parameter (or the mean) μ and the shape parameter (or the standard deviation) σ of the log-normal distribution can be expressed as follows (Stedinger and Cohn, 1986):

$$L(\mu,\sigma) = \prod_{i=1}^{s} f_X\left(x_i\right)\left\{\begin{bmatrix} h \\ k \end{bmatrix} F_X\left(X_0\right)^{h-k}\left[1 - F_X\left(X_0\right)\right]^k\right\} \prod_{j=1}^{k} f_Y\left(y_j\right) \tag{1}$$

where $f_X(x)$ and $F_X(x)$ are the probability density function and the cumulative distribution function of X, respectively. Also, as Y is always higher than X_0, $f_Y(y)$ can be expressed as follows:

$$f_Y(y) = \frac{f_X(y)}{1 - F_X\left(X_0\right)} \tag{2}$$

By substituting Equation (2) into Equation (1), the following likelihood function can be derived:

$$L(\mu,\sigma) = \prod_{i=1}^{s} f_X\left(x_i\right)\begin{bmatrix} h \\ k \end{bmatrix} F_X\left(X_0\right)^{h-k} \prod_{j=1}^{k} f_X\left(y_j\right) \tag{3}$$

Equation (3) is called the likelihood function of the censored data MLE. However, when the historical observations are available as a form of qualitative data, Equation (3) cannot be applied as it is. Equation (3) should be modified to consider only the frequency (i.e. the number of occurrences k) over the threshold value X_0 during the period of historical observation. These data are called the binomial censored data (Stedinger and Cohn, 1986). The likelihood function of the binomial censored data MLE can also be expressed by changing Equation (3) to consider the frequency of historical observations rather than their quantitative values. The model parameters μ and σ are selected as the set of values that maximize the likelihood function (Russell, 1982; Stedinger and Cohn, 1986):

$$L(\mu,\sigma) = \prod_{i=1}^{s} f_X\left(x_i\right)\left\{\begin{bmatrix} h \\ k \end{bmatrix} F_X\left(X_0\right)^{h-k}\left[1 - F_X\left(X_0\right)\right]^k\right\} \tag{4}$$

© 2017 Royal Meteorological Society

3. The Annals of the Choson Dynasty and the documentary records on rainfall events

3.1. Rainfall events in the Annals of the Choson Dynasty

The Annals of the Choson Dynasty are made up of 1894 books covering 472 years (1392–1863) of the history of the Choson Dynasty, from the reign of King Taejo, the founder, to the end of the reign of King Cheoljong (Figure 1). Additionally, there are two annals for the last two kings of the Choson Dynasty, King Gojong and King Sunjong. Since these two annals were published after the end of the Choson Dynasty, they are not officially included in the Annals. However, as the Annals generally refer to all the annals including the annals of the last two kings of the Choson Dynasty, this study also considered all the annals of the Choson Dynasty, from the founder King Taejo to the last King Sunjong, covering the entire period of 519 years from 1392 to 1910.

The Annals of the Choson Dynasty contain highly reliable records based on actual historical facts, compiled by historiographers who directly collected material, wrote and edited drafts, and then published the annals. These historiographers were also professional officials legally guaranteed independence in their record-keeping and the right to keep secrets. They needed to participate in and record all of the king's movements and all national affairs to create the Sacho ('Draft History'). Their daily drafts, various documents and daily records of the king and the government offices became the main sources for the compilation of the annals. Politics, economics, law, literature, diplomacy, military, industry, transportation, art, crafts and religion were among major issues in the annals. Disaster was also an important issue handled at that time, and major disasters could be recorded in the annals. When a king died and the coronation of his successor was completed, the annals of his reign were commenced by the Sillokcheong ('The Office for Annals Compilation'). The Sacho was not allowed to be read by anyone, including the king, and any historiographer who disclosed its contents was severely punished. These strict regulations lend great credibility to these records.

The preservation of the vast collection of the annals near the end of the Choson era in almost perfect condition is unparalleled in the world. The annals published for the previous king were made in four copies and stored with one set in Chunchugwan ('The Office of National History') and one set in each of three archives in deep mountainous sites chosen to avoid unforeseen damage and to ensure that the annals would be able to be passed on for posterity. This special care under a national system has made it possible today to discover the cultural and historical heritage of Korea recorded in these annals.

To broaden public access to the annals, the Korean government has supported the task of translating them into Korean from the original classical Chinese. After 26 years of effort, the Korean edition of the Annals of the Choson Dynasty was completed in 1993. To provide easy public access to the annals themselves and information about them, a CD-ROM version was made in 1995. This Korean edition of the Annals of the Choson Dynasty is now available from the website http://sillok.history.go.kr.

Many words were used to represent the rainfall events in the Choson Dynasty. At least eight different words were used to refer to rainfall, including 'tiny rain' (misty rain), 'thin rain' (drizzle), 'small rain' (light rain), 'falling rain' (rain), 'windy rain' (storm), 'sudden rain' (shower), 'big rain' (downpour) and 'explosive rain' (torrential rain). Also, many words were used to refer to floods. Among them, 'big water' (inundation), 'explosive flow' (torrential flow), 'ponding water' are frequently used. Additionally, when an inundation occurred due to 'big rain' or 'explosive rain', it was called a 'tremendous inundation'. Sometimes, when flood damage resulted from an unexpected sudden rain, the personal feelings of the historiographers influenced their word choice, and expressions such as 'bad water', 'ghost rain' (weird rain) and 'uncomfortable rain' (restricted narrow rain) were also used.

Kim *et al.* (2007a) studied the documentary records on the rainfall events in the Annals of the Choson Dynasty. Using various key words related to the rainfall events, they could identify a total of 556 documentary records related to the rainfall events. Yoo *et al.* (2015a) also repeated the same search with more keywords to derive a similar result. Figure 2 shows the annual variation of the number of documentary records. Interestingly, the records are not evenly distributed throughout the entire Choson Dynasty; about one half of the records are concentrated in the period of 140 years from the reign of King Myeongjong (1534) to that of King Hyeonjong (1674). It is uncertain whether the record is biased or whether the rainfall events were really that frequent during that period.

Recently, Yoo *et al.* (2015a) developed the rules of classification for these documents. The concept was simple, i.e. 'The magnitude of a rainfall event is strongly related to the size of the resulting disasters'. For example, for severer rainfall events, the frequency of the documents becomes higher and the length of the documents longer. The comments about the documents such as the size of the inundated area, the number of casualties and the amount of property loss provide important tips on the magnitude of the event. The size of the countermeasures is also assumed to be directly proportional to the magnitude of the rainfall events. Finally, the number of independent documents about the rainfall event was considered an important criterion. Table 1 shows the criteria applied to classify the rainfall events into catastrophic, extreme and severe events. The names of rainfall events, i.e. catastrophic, extreme and severe rainfall events, were subjectively given to represent worst, worse and bad events. At this moment, it is not easy to compare these rainfall events quantitatively, but some important quantitative differences could be derived as the analysis goes on. In particular, those that were referred to as an 'explosive rain' or 'explosive flow' were classified as catastrophic, along with those that produced landslides, a large number of drownings or a large number of houses drifting away. Additionally, the expression 'big rain like flow' was also assumed to be used to explain similar situations to those of 'explosive rain'.

Based on the rules of classification for the levels of rainfall events developed, a total of 326 events were identified as catastrophic, extreme or severe rainfall events during the 519 years of the Choson Dynasty (Table 2). Among them a total of 19 cases were classified as catastrophic rainfall events, 106 cases as extreme events and the remaining 201 cases were classified as severe rainfall events (Yoo *et al.*, 2015a). The chronological distributions of the catastrophic, extreme and severe rainfall events are compared in Figure 3. In this figure, the chronological distribution of the extreme rainfall events includes both the catastrophic and extreme events, and that of the severe rainfall events includes all of the catastrophic, extreme and severe events. Using these chronological distributions, one can check if the event occurs independently or not. The easiest way to check the independence of the occurrence of these events is to evaluate the inter-arrival times between consecutive events. If the distribution of the inter-arrival times follows an exponential distribution, it is generally assumed that the events occur independently following the Poisson process (Restrepo-Posada

© 2017 Royal Meteorological Society

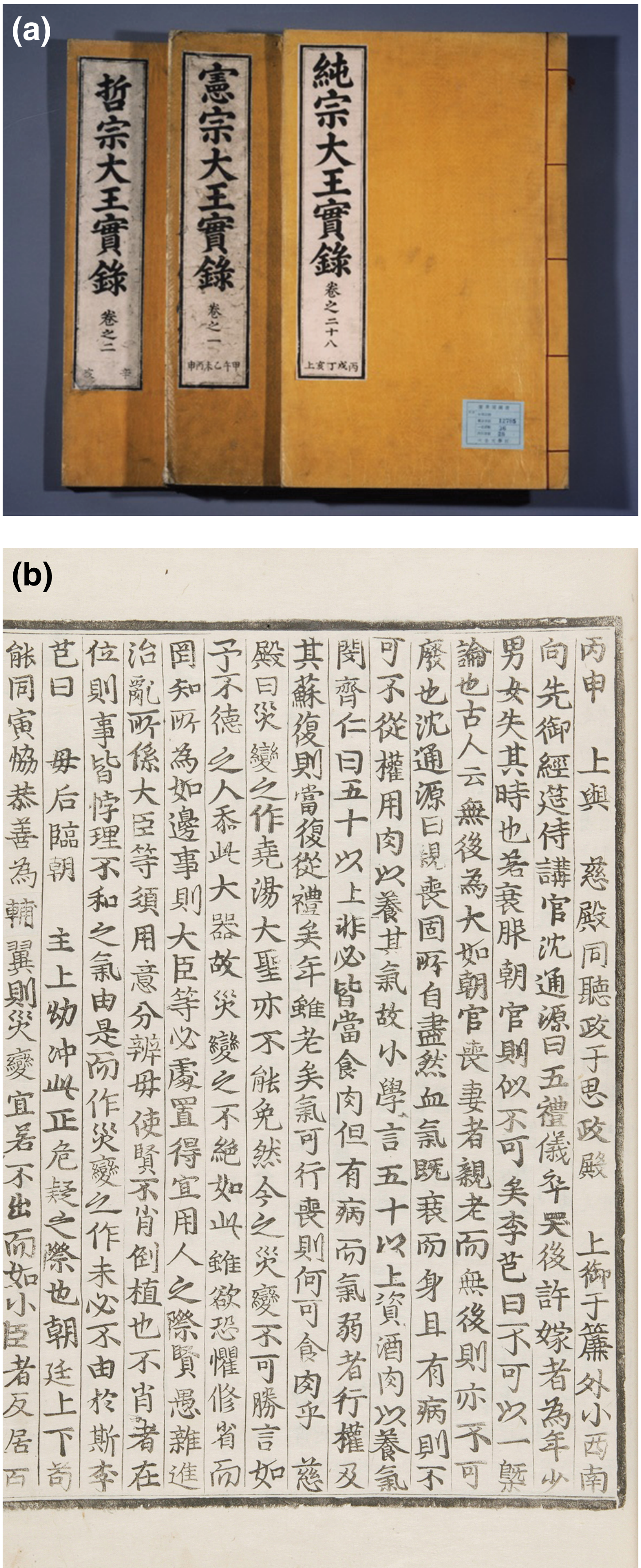

Figure 1. The Annals of the Choson Dynasty: (a) the three annals for Cheoljong, Hyeonjong and Sunjong, respectively; (b) part of the Annals of Myeongjong (Book 3, 11 June 1546).

© 2017 Royal Meteorological Society

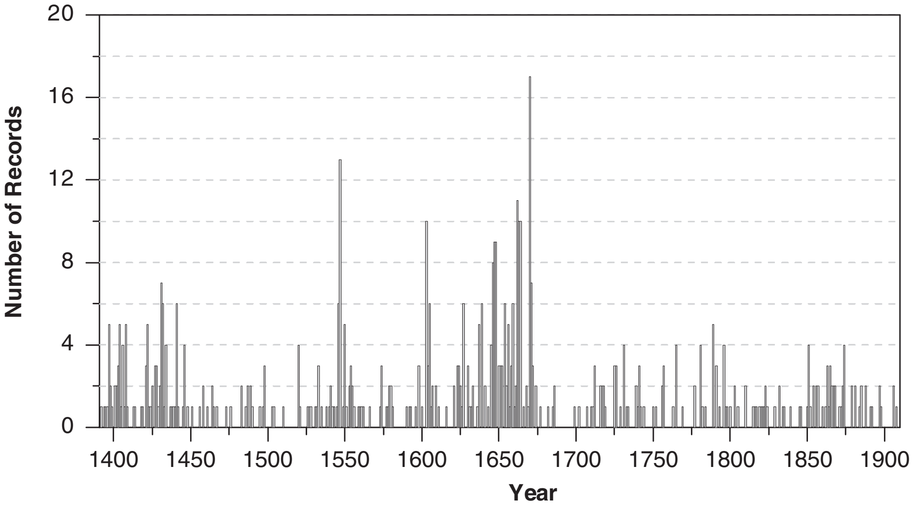

Figure 2. Annual variation of the number of documentary records in the Annals of the Choson Dynasty.

Table 1. Classification of rainfall events recorded in the Annals of the Choson Dynasty.

Grade	Terminology	Severities	Countermeasures	Number of document records
Catastrophic	• Explosive rain • Explosive flow	• Crushed people to death • Drowned people • Landslides • Sunken houses • Houses drifted away	• Dispatch of a relief official • Tax exemption • Relief money	~4
Extreme	• Big water • Big rain • Tremendous • Inundation • Bad water • Weird rain • Restricted narrow rain			2–3
Severe	• Big water • Big rain • Shower • Ponding water	• Sunken houses • Landslides	• Tax exemption • Holding ancestral rites	1–2

and Eagleson, 1982). Figure 4 shows three histograms of the inter-arrival times of the catastrophic, extreme and severe rainfall events, respectively, overlapped by exponential distributions. Both the χ^2 and the Kolmogorov–Smirnov goodness-of-fit tests with a significance level of 5% show that the occurrence of the catastrophic events and the extreme events was independent. However, the null hypothesis was rejected in the same tests for the severe rainfall events.

3.2. Chukwooki records

The Chukwooki, a traditional Korean rain gauge (Figure 5), was invented in 1441. It was used to measure rainfall in major cities in Korea, and the longest time period available in the data dates from 1777 for Seoul. Figure 6 shows the location of the Korean Peninsula and Seoul, the capital city of Korea. The location of the Chukwooki is very near to the location of the modern rain gauge station in Seoul, less than 1 km apart. The Chukwooki data could be recorded systematically on the king's order. It was known that both soil information and the Chukwooki rainfall data were used to estimate the amount of harvest. There have been many studies on the accuracy of the Chukwooki data in the 1990s (Lee and Jeong, 1992; Jung and Lim, 1994), and Jung (1999) finally showed that the Chukwooki data have a sufficient degree of accuracy for further analysis compared to modern measuring systems. Additionally, Jung *et al.* (2001) showed that the statistical characteristics of the Chukwooki time series (1783–1883) and modern time series (1911–1996) are very similar, including the diurnal cycle and its seasonal variation.

As investigated by Jung (1999), the Chukwooki data differ somewhat from modern measurement systems in several aspects such as the measuring units and the interval, operation schedule etc. The Chukwooki data were recorded with Korean foot-rules such as the Pun, Chi and Cha, which approximately correspond to 2, 20 and 200 mm, respectively; thus precipitation below about 2 mm (about 35–40 mm year^{-1}) could not be measured. Additionally, the Chukwooki records do not include the winter precipitation by snow (the precipitation amount from winter snow in Korea is about 40 mm year^{-1}). The sum of these two amounts is 75–80 mm year^{-1}, which is more or less the same as the difference between the Chukwooki and the modern data. Another aspect of the Chukwooki data to be considered is the data collection time interval. Unlike modern rainfall data, the data collection time interval was 2 h in the Choson Dynasty. This is due to the fact that the 12 h system was adopted in

© 2017 Royal Meteorological Society

Table 2. Number of catastrophic, extreme and severe rainfall events recorded in the Annals of the Choson Dynasty.

Periods	Numbers of storm events				Kings in each period
	Catastrophic	Extreme	Severe	Total	
1392–1500	5	15	23	43	Taejo, Jeongjong, Taejong, Sejong, Munjong, Danjong, Sejo, Yejong, Seongjong, Yeonsangun
1501–1600	4	13	29	46	Jungjong, Injong, Myeongjong, Seonjo
1601–1700	6	40	69	115	Gwanghaegun, Injo, Hyojong, Hyeonjong, Sukjong
1701–1800	1	17	37	55	Gyeongjong, Yeongjo, Jeongjo
1801–1910	3	21	43	67	Sunjo, Heonjong, Cheoljong, Gojong, Sunjong
Total	19	106	201	326	

the Choson Dynasty; thus, 1 h then is 2 h in the modern clock system.

The Chukwooki rainfall data is composed of the starting time, ending time and the total rainfall amount of an event, meaning that the duration and the total rainfall amount of a rainfall event were recorded (Jhun and Moon, 1997; Kim *et al*., 2007a). However, due to the limit of the Chukwooki size, in order to record the rainfall depth before the end of a day, sometimes a reading of the Chukwooki rainfall also had to be made. If the storm event was of particular interest such as the catastrophic events considered in this part of the study, more frequent and in-depth readings of the Chukwooki rainfall were also made.

Major rainfall events in the Annals of the Choson Dynasty were documented using words such as ‘big rain’, ‘explosive rain’ or ‘big water’. However, it is not particularly clear how these words were selected for a specific rainfall event. Fortunately, as the rainfall data measured by the Chukwooki are available during the period from 1777 to 1907, a one-step advanced evaluation of the documentary records could be possible (Yoo *et al*., 2015a). Comparison of the documentary records and the Chukwooki rainfall data was very helpful in verifying the rule for classifying the documentary records in the Annals of the Choson Dynasty.

3.3. Decision of threshold for consideration of catastrophic rainfall events

In Korea, systematic modern rainfall measurements since 1960 are available. Before then, the daily rainfall depth, along with sporadic 3 and 6 h rainfalls, had been recorded since 1907 (Korea Meteorological Administration, 2004). Before 1907 the Chukwooki was used to measure rainfall in the major cities in Korea. As mentioned in Section 3.2, the data structure of the Chukwooki rainfall is very basic, with the starting time, ending time, and the total rainfall depth of a rainfall event. Therefore, it was not possible to decide the rainfall duration consistently for all the data available. In this study, the 24 h rainfall amount was chosen as the most appropriate rainfall data for the frequency analysis since the annual maximum 24 h rainfall was assumed to represent the annual maximum rainfall event well (Yoo *et al*., 2015b). Rainfall of shorter duration was excluded due to the high level of uncertainty involved in the Chukwooki data. In addition, any rainfall longer than 24 h duration was also excluded due to the futility for practical use.

Figure 7 shows the annual maximum 24 h rainfall series since 1777. The highest value was recorded by the Chukwooki and the second highest value by a modern gauge. They are both higher than 400 mm. Also, the number of rainfall events of 300 mm or higher during the Chukwooki period is six, the same as that during the modern period. Also, Figure 8 compares the histograms of annual maximum 24 h rainfall series of the modern data and the Chukwooki data. The characteristics, as seen in the figure, are not very different from each other. Only the relative percentage of rainfall events of 200 mm or less looks slightly higher during the Chukwooki period, but it is not significant. Additionally, based on both the χ^2 and the Kolmogorov–Smirnov goodness-of-fit tests with a significance level of 5%, it could be confirmed that both annual maximum 24 h Chukwooki and modern data can be assumed to follow a log-normal distribution.

As described in Section 2, it is important to set a threshold value in order to consider the flood records in the frequency analysis. Fortunately, as the Chukwooki rainfall measurements are available for about 130 years before the end of the Choson Dynasty, the threshold value can be estimated with a sufficient degree of accuracy. In this study, only the catastrophic rainfall events in the Annals of the Choson Dynasty were considered in the frequency analysis. Among the rainfall events documented, these catastrophic rainfall events are believed to be of the highest degree of accuracy. A comparison of the rainfall events documented in the Annals and those recorded by the Chukwooki also shows that the classification results of these catastrophic rainfall events are the most accurate (Yoo *et al*., 2015a).

Only three catastrophic rainfall events were recorded during the 130 years of the Chukwooki data period at the end of the Choson Dynasty (see Figure 3). Table 3 summarizes the basic characteristics of these with their maximum rainfall depths depending on rainfall duration. It was found that all these three 24 h rainfall amounts were very similar to each other with a mean value of 388 mm. The amount of 388 mm for 24 h rainfall accounts for the top 1.5% of modern rainfall data, between the first and the second largest rainfall amounts in descending order. This mean value of the amount of 24 h rainfall for catastrophic rainfall events was used as the threshold value in the frequency analysis of this study.

4. Results of frequency analysis

4.1. Parameters and design rainfall estimated

As mentioned before, the parameters of the log-normal distribution were estimated by the binomial censored data MLE. The results are shown in Table 4. As can be seen, the parameter estimation was repeated many times to see the effect of considering different types of data as well as data length. Summarizing the results, first the effect of considering different types of data is obvious. Additional consideration of the data between 1911 and 1960 decreased the location parameter while increasing the shape parameter. On the other hand, additional consideration of the catastrophic rainfall events during the Chukwooki data period (1777–1907) decreased both the location and the shape parameters. Finally, additional consideration of the

© 2017 Royal Meteorological Society

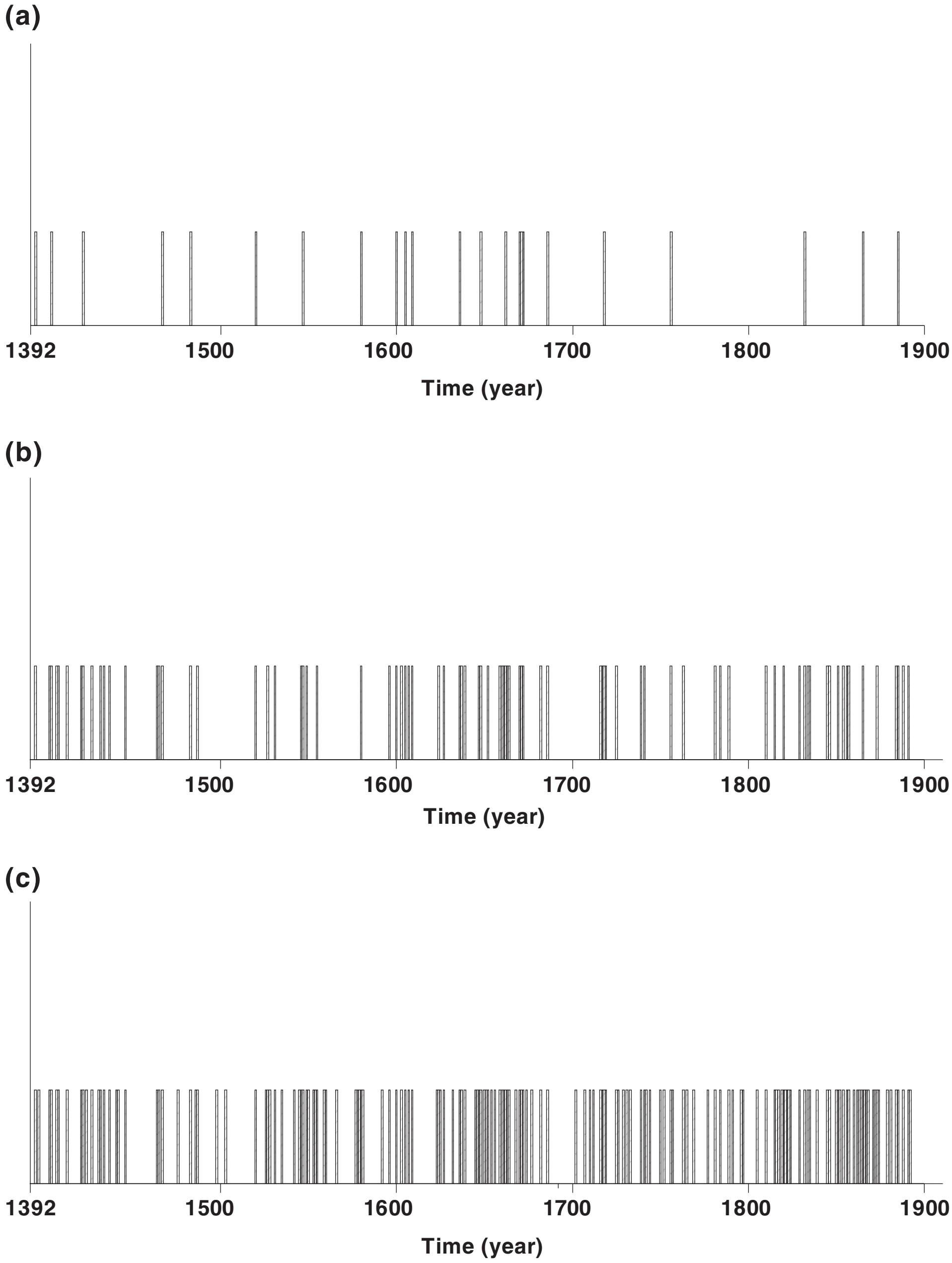

Figure 3. Chronological occurrence of the catastrophic, extreme and severe rainfall events: (a) catastrophic storm events; (b) catastrophic plus extreme storm events; (c) catastrophic plus extreme plus severe storm events.

catastrophic rainfall events recorded in the Annals of the Choson Dynasty (1392–1776) increased both the location and the shape parameters.

A long-term trend in the parameters estimated is also noticeable. The presence of an inflection point around 1700 is of special interest. Both the location parameter and the shape parameter were found to be smallest, 4.89 and 0.46 respectively. This characteristic behaviour of the estimated parameters also affects the design rainfalls significantly (Figure 9). As can be seen in Figure 9, the estimated 24 h design rainfalls all show minima around 1700. The 100 year 24 h rainfall amount estimated was less than 400 mm around 1700 but increased to 446.1 mm in 1392. This amount is very similar to that estimated by considering only the modern rainfall data since 1961. Similarly, the 500 year 24 h rainfall amount estimated was slightly less than 500 mm around 1700 but increased to 588.5 mm in 1392. This amount is also very similar to that estimated by considering only the modern rainfall data since 1961.

It is also interesting to check if there is any effect of considering three different datasets on the results of frequency analysis. Basically, the years 1910 and 1777 could be breakpoints in the series, when the catastrophic rainfall events during the Chukwooki data period and those in the Annals of the Choson Dynasty began to be considered. However, the effect of merging three different datasets was found negligible and a long-term consistent behaviour could be derived. The effect of adding the catastrophic rainfall events during the Chukwooki data period did not change the trend in the results of the frequency analysis at all. This was also the same when adding the catastrophic rainfall events in the Annals of the Choson Dynasty. Overall, the occurrences of catastrophic events seem to be consistent in all data collection periods of the modern data, Chukwooki data and documentary records.

© 2017 Royal Meteorological Society

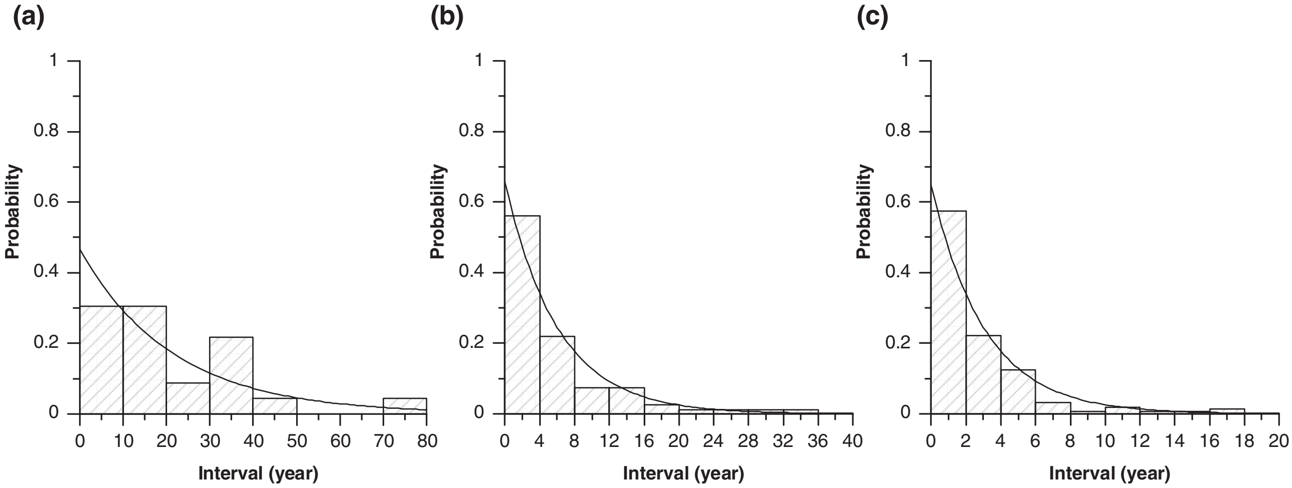

Figure 4. Goodness-of-fit tests for the inter-arrival times of the catastrophic (a), extreme (b) and severe (c) rainfall events.

4.2. Uncertainty in the design rainfalls

In this part of the study, the uncertainty in the estimated design rainfalls due to the possible uncertainty in the threshold value and the number of catastrophic rainfall events was examined. The results summarized in Section 4.1 are those conditioned on the threshold value of the 24 h rainfall amount, 388 mm, and the number of catastrophic rainfall events, 19, including the three in the Chukwooki data.

First, the possible amounts of uncertainty in the threshold value were assumed to be 10 mm (about 2%). Here, 10 mm is the maximum difference between the 24 h rainfall amounts of three catastrophic rainfall events recorded by the Chukwooki. For the two boundary values of the threshold selected by considering the uncertainty amount, i.e. 378 and 398 mm, the frequency analysis was repeated to get the design rainfalls (Table 5). It was found that, for short return periods such as 5 and 20 years, the change in the design rainfalls estimated was just 1–4 mm. For long return periods such as 100 and 500 years, the change in the design rainfalls increased to 12 mm. These changes are just about 1% for the short return periods and about 2% for the long return periods. Additionally, as an extreme case to assume the uncertainty level to be 25%, i.e. from 300 to 500 mm, the change in the design rainfalls was estimated to be far less than 20% even for long return periods. Overall, it can be concluded that the threshold value is not very sensitive to the probable rainfall and the uncertainty in the design rainfalls estimated for the 388 mm threshold is quite low.

Second, the possible uncertainty in the number of catastrophic rainfall events was assumed to be about 10%. This uncertainty level of 10% was arbitrarily introduced for the sensitivity analysis. Considering a 10% level of uncertainty, the number of catastrophic rainfall events could be from 17 to 21, excluding three events in the Chukwooki data. For the two extreme boundary values, 17 and 21, the frequency analysis was repeated to get the design rainfalls (Table 6). It was found that, for short return periods like 5 and 20 years, the change in the estimated design rainfalls was just 3–5 mm. For long return periods like 100 and 500 years, the change in the design rainfalls increased to 16 mm. These changes are just about 2% for the short return periods and about 3% for the long return periods. Overall, it can be concluded that the sensitivity of the design rainfalls to the number of catastrophic rainfall events is very low. As an extreme case, assuming the uncertainty level to be 30%, i.e. the number of catastrophic rainfall events was 13–23, the change in the estimated design rainfalls, even for long return periods, was found to be far less than 10%. Similar to the previous case of a threshold value applied to the frequency analysis, the uncertainty in the design rainfalls estimated for the number of catastrophic rainfall events 19 was found to be very low.

4.3. Confidence intervals of design rainfalls

The confidence intervals of the estimated design rainfalls are given in Table 7. These confidence intervals are those estimated by considering the confidence interval of the location parameter only. Because the data used in the frequency analysis were those reconstructed from the Annals of the Choson Dynasty, it was difficult to apply a confidence interval to judge the data themselves. Instead, a confidence interval could be applied to only the mean values (distributional location parameters) as secondary information. Given this characteristic, the confidence interval for the location parameter of the log-normal distribution was estimated in this study and used for deriving the confidence intervals of the design rainfalls.

The confidence interval of the location parameter can be derived by considering the accuracy level of the data used. With the significance level given, the confidence interval (CI) can be derived by using the binomial distribution (Gelman *et al*., 1997; Agresti and Coull, 1998; Brown *et al*., 2001; Newcombe, 2001). That is:

$$\mathrm{CI} = z_{1-\alpha/2}\left[p\left(1-p\right)/n\right]^{-1/2} \quad (5)$$

where p is the accuracy level given for the data, n is the number of data points and $z_{1-\alpha/2}$ is the standard normal random variable at a significance level α. In particular, $z_{1-\alpha/2}$ becomes 1.96 at a significance level of 5%.

As can be seen in Table 7, the confidence intervals of the design rainfalls are extremely dependent on the accuracy level of the data used for the frequency analysis. When the accuracy level is 99% (and at a significance level of 5%), the range between the lower and upper bounds of the 5 year design rainfall is just 16.5 mm, which becomes 34.7 mm for the 100 year design rainfall and about 45.8 mm for the 500 year design rainfall. On the other hand, with an accuracy level of 50%, the range between the lower and upper bounds of the 5 year design rainfall becomes about 83 mm, which turns into 175 mm for the 100 year design rainfall and about 231 mm for the 500 year design rainfall.

© 2017 Royal Meteorological Society

Figure 5. Chukwooki, an old Korean rain gauge used in the Choson Dynasty (the above Chukwooki was known to be made around 1770 in Korea).

It is also interesting to review the effect of data length on the confidence interval. Table 8 summarizes the confidence intervals for two accuracy levels of 0.99 and 0.50. Where the accuracy level of the data is 99%, the confidence interval of the mean value of the 50 year long data is ±2.88%. When the data length is 101 years, the confidence interval becomes ±1.99%, and it becomes less than ±1.30% when the length of the data increases to 234 years (the length of both modern data and the Chukwooki data). When the length of the data increases to 619 years (the length of modern data, the Chukwooki data and the records in the Annals of the Choson Dynasty altogether), the confidence interval becomes just ±0.79%. That is, as the data length increases, the uncertainty involved in the estimates of the design rainfall becomes much smaller. Simply considering the records of the rainfall events in the Annals of the Choson Dynasty decreases the uncertainty to almost a half. The result is very appealing even when the accuracy level of the data is very low. Assume that the documents on the rainfall events in the Annals of the Choson Dynasty are very inaccurate or exaggerated, to give an accuracy level of the data of just 50%. Even in this case, the confidence interval becomes less than ±4%. This confidence interval of ±4% is that which can be obtained from the frequency analysis when the length of the data is more than 30 years and the accuracy of the data is higher than 99%.

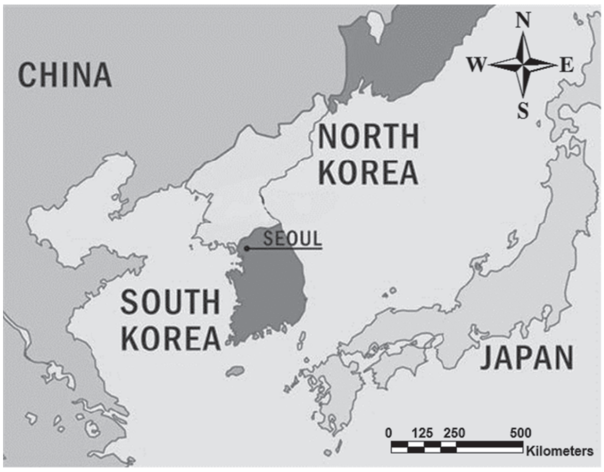

Figure 6. Location of the Korean Peninsula and Seoul, the capital city of Korea.

5. Conclusions

In this study, the flood records in the Annals of the Choson Dynasty (1392–1910) were reviewed and considered for the frequency analysis of rainfall in Seoul, Korea. Based on the rules of classification for the levels of rainfall events developed, a total of 326 events was identified as catastrophic, extreme or severe rainfall events during the 519 years of the Choson Dynasty. A comparison of the rainfall events documented in the Annals of the Choson Dynasty and the data (1777–1907) from the Chukwooki (an old Korean rain gauge) showed that the catastrophic events were most accurate, and they were then considered in the frequency analysis. The Chukwooki data were also used to derive the threshold value for considering the catastrophic events in the Annals of the Choson Dynasty. Finally, the design rainfalls were estimated using the binomial censored data maximum likelihood estimator. Additionally, the sensitivity of the design rainfalls to

Table 3. Basic characteristics of three catastrophic rainfall events included in both the Chukwooki data and the Annals of the Choson Dynasty.

Year	Storm duration (h)	Total rainfall depth (mm)	Mean rainfall intensity (mm h^{-1})	Maximum rainfall depth (mm)		
				6 h	12 h	24 h
1832	88	516	6	124	247	386
1865	24	385	16	120	240	385
1885	96	566	6	147	294	392
Mean				130	260	388

© 2017 Royal Meteorological Society

Table 4. Parameters estimated by increasing the data length.

Periods	Data length (years)	Number of storm events	Parameters		Remarks
			Location	Shape	
2010 to 1961	50	—	5.07	0.43	Modern data
2010 to 1910	101	—	4.91	0.47	
2010 to 1850	161	2	4.88	0.46	Both Chukwooki data and documentary records available
2010 to 1800	211	3	4.90	0.46	
2010 to 1777	234	3	4.90	0.45	
2010 to 1700	311	4	4.89	0.46	Only documentary records available
2010 to 1650	361	6	4.90	0.46	
2010 to 1600	411	10	4.91	0.48	
2010 to 1550	461	11	4.91	0.48	
2010 to 1500	511	14	4.92	0.49	
2010 to 1450	561	16	4.93	0.50	
2010 to 1392	619	19	4.93	0.50	

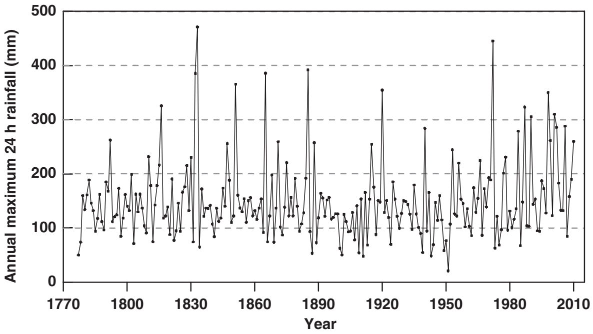

Figure 7. Annual maximum 24 h rainfall series at Seoul, Korea (the Chukwooki data were provided by Dr H. Kim of Korea Institute of Construction Technology and the modern data by Korea Meteorological Administration).

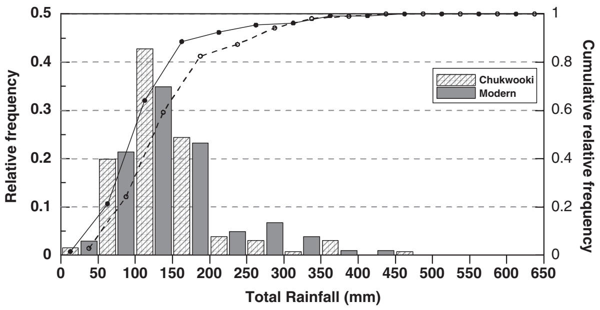

Figure 8. Histograms and cumulative relative frequency curves of annual maximum 24 h rainfall series of the modern and Chukwooki data.

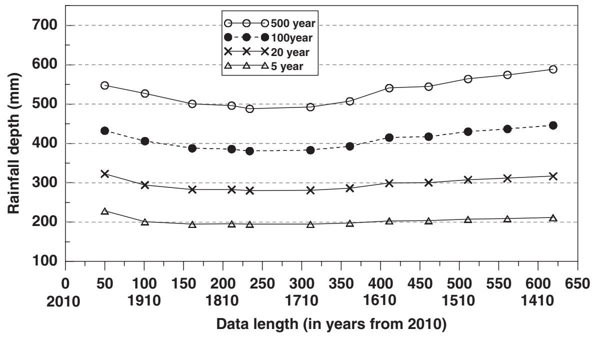

Figure 9. Change of design rainfall with respect to the data length and the return period.

the threshold value, the number of flood records considered, and the accuracy level of the flood records was also analysed to evaluate the quality of the frequency analysis. The results can be summarized as follows.

First, the frequency analysis was done for 24 h rainfall amounts as it was assumed to represent the three different rainfall records well (the documentary records in the Annals of the Choson Dynasty, the Chukwooki rain gauge data and the modern rain gauge data). Also, to secure the highest degree of accuracy for the classification results of the documentary records, only the catastrophic events (the highest level of flood) were considered in the frequency analysis. A total number of 19 were classified as catastrophic events in the Annals of the Choson Dynasty, among which three were observed during the period of the Chukwooki records. The Chukwooki data for those three events were used to decide the threshold value of the frequency analysis, which was 388 mm.

Second, the sensitivity analysis of the design rainfall was done with respect to the threshold value, the number of catastrophic events and the accuracy level of the flood records in the Annals of the Choson Dynasty. The possible range of the threshold value was selected by considering the range of the Chukwooki data for three catastrophic events while the uncertainty range of the number of catastrophic events and the accuracy level of the flood records were assumed arbitrarily to be about 10%. As a result, the sensitivities of the probable rainfall on these three factors were all found to be smaller than the uncertainty level considered. Even in an extreme case of considering a very low accuracy level of the flood records, the uncertainty of the probable rainfall was found to be much smaller than that of the accuracy level. Consequently, it was found that the length of the flood records in the Annals of the Choson Dynasty is long enough to overcome the possible uncertainty of the flood records.

Finally, the design rainfalls estimated considering the flood records in the Annals of the Choson Dynasty were found to be very similar to those estimated by considering only the modern data. It was also found that the design rainfalls were steadily decreasing on increasing the record length, becoming smallest around 1700, and increasing again. However, the confidence interval of the design rainfalls was found to decrease steadily as the record length increased. The confidence intervals estimated

Table 5. Variation of design rainfall depending on the threshold values considered.

Thresholds (mm)	Parameters		Return periods (year)					
	Location	Shape	5	20	50	100	200	500
500	4.96	0.52	222.0	337.7	418.1	482.1	549.1	643.0
400	4.94	0.51	213.4	320.9	395.1	453.7	515.0	600.5
388	4.93	0.50	211.6	316.8	389.0	446.1	505.6	588.5
380	4.93	0.50	210.5	314.3	385.5	441.7	500.2	581.7
300	4.89	0.46	196.2	283.4	341.8	387.3	434.2	498.6

Table 6. Variation of design rainfall depending on the total numbers of catastrophic rainfall events considered.

Number of events	Parameters		Return periods (year)					
	Location	Shape	5	20	50	100	200	500
13	4.91	0.48	203.4	299.2	364.1	415.0	467.8	540.9
17	4.93	0.50	208.9	311.0	380.8	435.8	493.1	572.7
19	4.93	0.50	211.6	316.8	389.0	446.1	505.6	588.5
21	4.94	0.51	214.2	322.4	397.1	456.2	517.9	604.1
23	4.94	0.52	216.7	328.1	405.2	466.4	530.5	620.1

by considering all the records available were found to be just one third of those of the modern data only.

Thus, it may be unreasonable to consider that the confidence interval is an absolute index for the accuracy of estimates of design rainfalls. However, it was confirmed that the impact of reconstructing long-term historical data is valuable even when the quality of the historical data is not very high compared with modern measurements.

Acknowledgements

This research was supported by the Basic Science Research Program through the Korea Research Foundation (KRF-2008-313-D01083) and the National Research Foundation of Korea (NRF) through the Ministry of Education, Science and Technology (no. 2010-0014566). All contributions are gratefully acknowledged.

Table 7. 95% confidence intervals of design rainfall derived for some return periods and accuracy levels of the data.

Data accuracy	Confidence interval	Parameters		Return periods (year)					
		Location	Shape	5	20	50	100	200	500
Estimates		4.93	0.50	211.6	316.8	389.0	446.1	505.6	588.5
99%	Upper	4.97	—	220.1	329.4	404.5	463.8	525.7	611.8
	Lower	4.89	—	203.6	304.7	374.2	429.1	486.3	566.0
50%	Upper	5.13	—	257.3	385.2	473.0	542.3	614.7	715.4
	Lower	4.74	—	174.1	260.6	320.0	366.9	415.9	484.0

Table 8. Confidence intervals of design rainfall derived depending on the accuracy level *p* and data length.

Periods	Data length (years)	Parameters		Confidence intervals of location parameter (%)	
		Location	Shape	$p = 0.99$	$p = 0.50$
2010–1961	50	5.07	0.43	±2.88%	±14.45%
2010–1910	101	4.91	0.47	±1.99%	±10.00%
2010–1850	161	4.88	0.46	±1.56%	±7.85%
2010–1800	211	4.90	0.46	±1.36%	±6.83%
2010–1777	234	4.90	0.45	±1.30%	±6.52%
2010–1700	311	4.89	0.46	±1.12%	±5.63%
2010–1650	361	4.90	0.46	±1.04%	±5.22%
2010–1600	411	4.91	0.48	±0.97%	±4.88%
2010–1550	461	4.91	0.48	±0.92%	±4.60%
2010–1500	511	4.92	0.49	±0.87%	±4.37%
2010–1450	561	4.93	0.50	±0.83%	±4.17%
2010–1392	619	4.93	0.50	±0.79%	±3.96%

© 2017 Royal Meteorological Society

References

Agresti A, Coull BA. 1998. Approximate is better than 'exact' for interval estimation of binomial proportions. *Am. Stat.* **52**: 119–126.

Bishop YM, Fienberg SE, Holland PW. 1975. *Discrete Multivariate Analysis: Theory and Practice*. MIT Press: Cambridge, MA.

Brown LD, Cai TT, DasGupta A. 2001. Interval estimation for a binomial proportion (with discussion). *Stat. Sci.* **16**: 101–133.

Cohen AC. 1950. Estimating the mean and variance of normal populations from singly truncated and doubly truncated samples. *Ann. Math. Stat.* **21**(4): 557–569.

Cohen AC. 1976. Progressively censored sampling in the three parameter log-normal distribution. *Technometrics* **18**(1): 99–103.

Condie R, Lee KA. 1982. Flood frequency analysis with historic information. *J. Hydrol.* **58**: 47–61.

Condie R, Pilon PJ. 1983. Fitting the log-Pearson type III distribution to censored samples: an application to flood frequency analysis with historic information. In *Proceedings of the 6th Canadian Hydrotechnical Conference*. Canadian Society for Civil Engineering: Ottawa; 11–22.

Frances F. 1998. Using the TCEV distribution function with systematic and non-systematic data in a regional flood frequency analysis. *Stoch. Hydrol. Hydraul.* **12**(4): 267–283.

Gelman A, Carlin JB, Stern HS, Rubin DB. 1997. *Bayesian Data Analysis*. Chapman and Hall: London.

Hald A. 1949. Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point. *Skand. Aktuartidskr.* **32**: 119–134.

Hwang SH, Kim JH, Yoo C, Lee JH. 2010. Characteristics of inter-monthly climatic change appeared in long-term Seoul rainfall. *J. Korean Soc. Civil Eng.* **30**(1B): 1–11.

Jhun JG, Moon BK. 1997. Restorations and analyses of rainfall amount observed by Chukwooki. *Asia-Pac. J. Atmos. Sci.* **33**(4): 691–707.

Jung HS. 1999. Interpretation of the transient variation in the time series of precipitation amounts in Seoul, PhD thesis, Seoul National University, Seoul.

Jung HS, Lim GH. 1994. On the monthly precipitation amounts and number of precipitation days in Seoul, 1770–1907. *Asia-Pac. J. Atmos. Sci.* **30**(4): 487–505.

Jung HS, Lim GH, Oh JH. 2001. Interpretation of the transient variations in the time series of precipitation amounts in Seoul, Korea: Part 1. Diurnal variation. *J. Clim.* **14**(13): 2989–3004.

Kim HJ. 2000. Flood investigation in the Annals of Choson Dynasty. *J. Korea Water Resour. Assoc.* **33**(1): 365–370.

Kim D, Yoo C, Kim HJ. 2007a. Evaluation of major storm events both measured by Chukwooki and recorded in Annals of Chosen Dynasty: 2. Quantitative approach. *J. Korea Water Resour. Assoc.* **40**(7): 545–554.

Kim K, Yoo C, Park M, Kim HJ. 2007b. Evaluation for usefulness of Chukwookee data in rainfall frequency analysis. *J. Korea Water Resour. Assoc.* **40**(11): 851–859.

Kirby W. 1981. Instructions for peak flow file. US Geological Survey Open File Report 79-1336-1. United States Geological Survey: VA.

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

Kuczera G. 1999. Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference. *Water Resour. Res.* **35**(5): 1551–1557.

Lee DR, Jeong SM. 1992. A spatial-temporal characteristic of rainfall in the Han river basins. *J. Korean Assoc. Hydrol. Sci.* **25**(4): 75–85.

Leese MN. 1973. Use of censored data in the estimation of Gumbel distribution parameters for annual maximum flood series. *Water Resour. Res.* **9**(6): 1534–1542.

Liu KB, Shen C, Louie KS. 2001. A 1000-year history of typhoon landfalls in Guangdong, southern China, reconstructed from Chinese historical documentary records. *Ann. Assoc. Am. Geogr.* **91**(3): 453–464.

Newcombe RG. 2001. Logit confidence intervals and the inverse sinh transformation. *Am. Stat.* **55**: 200–202.

Park M. 2009. Probabilistic occurrence characteristics and climatic variabilities of extreme storm events, PhD thesis, Korea University, Seoul.

Park M, Yoo C, Kim H, Jun C. 2014. Bivariate frequency analysis of annual maximum storm event series in Seoul, Korea. *J. Hydrol. Eng.* **19**(6): 1080–1088.

Reed D. 1999. *Flood Estimation Handbook: Overview* (Vol.1). Institute of Hydrology: Wallingford.

Reis DS Jr, Stedinger JR. 2005. Bayesian MCMC flood frequency analysis with historical information. *J. Hydrol.* **313**: 97–116.

Restrepo-Posada PJ, Eagleson PS. 1982. Identification of independent rainstorms. *J. Hydrol.* **55**: 303–319.

Russell SO. 1982. Flood probability estimation. *J. Hydraul. Div.* **108**(1): 63–73.

Sivapalan M, Takeuchi K, Franks SW, Gupta VK, Karambiri H, Lakshmi V, et al. 2003. IAHS decade on predictions in ungauged basins (PUB), 2003–2012: shaping an exciting future for the hydrological sciences. *Hydrol. Sci. J.* **48**(6): 857–880.

Stedinger JR, Cohn TA. 1986. Flood frequency analysis with historical and palaeoflood information. *Water Resour. Res.* **22**(5): 785–793.

Stedinger JR, Lu L-H. 1995. Appraisal of regional and index flood quantile estimators. *Stoch. Hydrol. Hydraul.* **9**(1): 49–75.

Tasker GD, Thomas WO Jr. 1978. Flood-frequency analyses with prerecord information. *J. Hydraul. Div.* **104**(2): 249–259.

United States Water Resources Council Hydrology Committee (USWRC). 1967. *Guidelines for Determining Flood Flow Frequency*. Bulletin 15. US Government Printing Office: Washington, DC.

United States Water Resources Council Hydrology Committee (USWRC). 1982. *Guidelines for Determining Flood Flow Frequency*. Bulletin 17B. US Government Printing Office: Washington, DC.

Wang B, Ding Q, Jhun JG. 2006. Trends in Seoul (1778–2004) summer precipitation. *Geophys. Res. Lett.* **33**: L15803.

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

Yoo C, Park M, Kim HJ, Choi J, Sin J, Jun C. 2015a. Classification of the documentary-recorded storm events in the Annals of the Choson Dynasty (1392–1910), Korea. *J. Hydrol.* **520**: 387–396.

Yoo C, Park M, Kim HJ, Jun C. 2015b. Comparison of rainfall event data measured by old Korean rain gauge (Chukwooki) and modern rain gauge in Seoul, Korea. *J. Clim.* (under review) .

© 2017 Royal Meteorological Society