water

MDPI

Article

Effect of Multicollinearity on the Bivariate Frequency Analysis of Annual Maximum Rainfall Events

Chulsang Yoo and Eunsaem Cho *

School of Civil, Environmental and Architectural Engineering, College of Engineering, Korea University, Seoul 02841, Korea; envchul@korea.ac.kr

* Correspondence: saemrnt@gmail.com; Tel.: +82-10-4547-2681

Received: 21 March 2019; Accepted: 26 April 2019; Published: 29 April 2019

Abstract: A rainfall event, simplified by a rectangular pulse, is defined by three components: the rainfall duration, the total rainfall depth, and mean rainfall intensity. However, as the mean rainfall intensity can be calculated by the total rainfall depth divided by the rainfall duration, any two components can fully define the rainfall event (i.e., one component must be redundant). The frequency analysis of a rainfall event also considers just two components selected rather arbitrarily out of these three components. However, this study argues that the two components should be selected properly or the result of frequency analysis can be significantly biased. This study fully discusses this selection problem with the annual maximum rainfall events from Seoul, Korea. In fact, this issue is closely related with the multicollinearity in the multivariate regression analysis, which indicates that as interdependency among variables grows the variance of the regression coefficient also increases to result in the low quality of resulting estimate. The findings of this study are summarized as follows: (1) The results of frequency analysis are totally different according to the selected two variables out of three. (2) Among three results, the result considering the total rainfall depth and the mean rainfall intensity is found to be the most reasonable. (3) This result is fully supported by the multicollinearity issue among the correlated variables. The rainfall duration should be excluded in the frequency analysis of a rainfall event as its variance inflation factor is very high.

Keywords: annual maximum rainfall event; bivariate frequency analysis; copula; multicollinearity

1. Introduction

The bivariate frequency analysis (BFA) is used for the quantification of the probabilistic characteristic of two correlated variables. This analysis has been conducted in a wide range of study areas. De Haan and De Ronde [1] evaluated the risk of wave overtopping by conducting the BFA on the still water level and the wave height. Yue [2] assessed the applicability of the bivariate exponential distribution for two univariate data that follow the exponential distribution. Yue and Wang [3] compared the Gumbel mixed model and the Gumbel logistic model, which are used for the BFA.

In terms of hydrology and water resources, the BFA has been conducted mostly for processes like the rainfall event, drought, and river discharge. In the case of the rainfall event, two variables among several characteristics, such as the mean rainfall intensity, rainfall duration, total rainfall depth, and maximum rainfall intensity, are selected to perform the analysis [4,5]. For the drought, two variables among the characteristics of the drought event, such as the average drought severity, drought duration, total drought severity, and maximum drought severity, are usually used for the analysis [6–8]. The analysis of the river discharge is also similar [9–11].

Bivariate probability distribution functions (PDFs) have been generally used for the BFA. The bivariate exponential distribution, bivariate gamma distribution, and bivariate extremal distribution have been frequently used regarding hydrology and water resources [12–15]. The use of the bivariate probability distribution, such as bivariate exponential distribution [16] and bivariate gamma distribution [17], is only valid when two variables follow the same PDF. Unfortunately, this assumption is not valid in most cases of BFA. For example, the PDF for the mean rainfall intensity of a rainfall event is different from that for rainfall duration. The former is generally modeled by the log-normal distribution, but the exponential distribution is used for the latter [18]. In this case, the copula can be a sound option [19] and many examples where the copula is applied can also be found in hydrology and water resources [7,12,18,20]. De Michele et al. [12] analyzed the flood peak and flood volume data from Northern Italy using bivariate copulas to check adequacy of dam spillway. Zhang and Singh [18] derived bivariate distribution of rainfall variables using copula and tested the applicability on the data from the Amite river basin in Louisiana, United States. Mirabbasi et al. [7] applied bivariate copulas on drought duration and severity data from Iran and conducted drought frequency analysis.

The objective of this study is to provide a guideline about the selection of two variables for the BFA of rainfall events. 'Rainfall event' is a basic unit for characterizing storm. In the case of 'independent rainfall event' is defined as a rainfall event which is not correlated with any other rainfall events. It is possible to define interdependent rainfall event by separating rainfall events with correlation time, which means the minimum time required to make rainfall event independent. This correlation time is also referred to Inter-event Time Definition (IETD).

In Korea, 10 hours is generally applied as the IETD [21,22]. However, this study considered various IETD based on frequency of independent rainfall event to determine reasonable IETD. As a result, the 12 hours of IETD was effective to separate independent rainfall event. This IETD was also considered in this study to separate the independent rainfall events. Before separating the rainfall event, a threshold value was applied to distinguish the rain/no rain condition. In this study, 1 mm/hour was considered for this purpose.

It is generally accepted that the structure of a rainfall event can be simplified with the rectangular pulse model [23–27]. In the rectangular pulse model, a rainfall event is quantified by three components; the rainfall duration, the total rainfall depth and mean rainfall intensity. Here, the mean rainfall intensity can be calculated by the total rainfall depth divided by the rainfall duration. Thus, any two components can fully define the rainfall event and one component must be redundant (Figure 1).

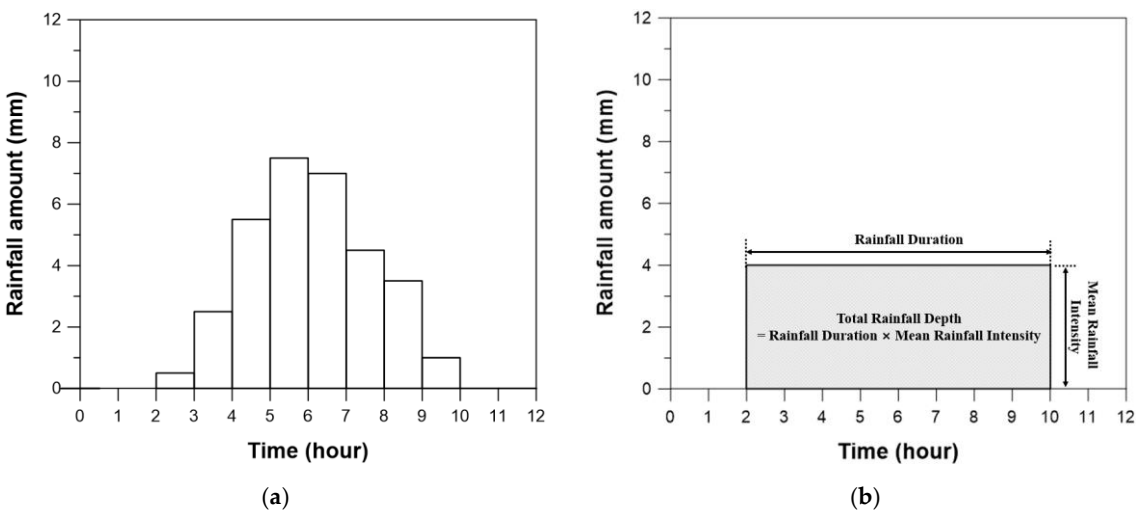


Figure 1. Simplification of an observed rainfall event (**a**) by a rectangular pulse (**b**).

However, a way to properly select the two components of a rainfall event has not been established yet and the way that the random selection of two variables may affect the results of the

BFA has not been confirmed. Therefore, many researchers had performed the BFA by selecting two variables through their subjective decision [13,20,28,29]. For example, Fu and Butler [20] selected the rainfall duration and the total rainfall depth of a rainfall event to perform the BFA. Jain et al. [29] selected the rainfall duration and the mean rainfall intensity for the same BFA. Park et al. [13] used the total rainfall depth and the mean rainfall intensity for the BFA.

In fact, the authors assume that the selection issue of the two variables for the BFA is closely related with the multicollinearity issue in regression analysis. The dependency among predictors, which is generally called interdependency, deteriorates the quality of estimate result [30]. Thus, the two variables should be selected to minimize this problem. This manuscript is composed of a total of five sections including introduction and conclusions. Theoretical background of the multicollinearity problem and the BFA using copula is summarized in Section 2. As an application, the annual maximum rainfall events observed in Seoul, Korea are analyzed in Sections 3 and 4. Finally, the conclusions and discussion are given in Section 5.

2. Theoretical Background

2.1. Multicollinearity

2.1.1. Multicollinearity Problem in Regression Analysis

Regression analysis is conducted to find a relation between two or more variables, of which at least one is subject to random variation. For example, when a linear model is concerned between the predictors X and response Y, the least square method leads to the estimates

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{1}$$

and the variance-covariance matrix of the estimates is given as follows [31]

$$\mathrm{var}(\hat{\beta}) = \sigma^2 (X^T X)^{-1} \tag{2}$$

Multicollinearity is assumed here as an interdependency condition in X. Multicollinearity can thus be defined as a lack of independence or a presence of interdependence. It is generally evaluated by the correlation matrix $(X^T X)$. As interdependence among predictors grows, the correlation matrix $(X^T X)$ approaches singularity and elements of the inverse matrix $(X^T X)$ become infinite. As an extreme case, when perfect linear dependence within a predictor set is secured, a perfect singularity on $(X^T X)$ is resulted and lead a completely indeterminate set of parameter estimates $\hat{\beta}$. As elements of the inverse matrix $(X^T X)$ becomes infinite, variances of regression coefficients also become infinite and the variance can be allocated completely arbitrarily among predictors [32,33]. Similarly, the large variances on regression coefficients produced by interdependent variables can lead the low quality of resulting parameter estimates [33,34]. It becomes almost impossible to distinguish the independent contribution of a predictor to the variance of the response.

In the process of analyzing several variables, the multicollinearity is also an important factor that distorts the result. Highly correlated variables are the main cause of the multicollinearity problem. It is known that the effect of the multicollinearity distortion increases with the increasing of the degree of correlation [32,35–37]. If highly correlated variables are included in the multilinear regression equation, the predicted result can be distorted due to the multicollinearity problem of these variables. This problem is generally evaluated through a quantification of the variance inflation from a variable [38–40]. The variance inflation factor is defined as a measure to quantify the level of the multicollinearity in a multiple regression analysis. The variance inflation factor can be calculated as follows:

$$VIF_i = \frac{1}{1 - R_i^2} \tag{3}$$

where R_i is the coefficient of determination that indicates how much the variable can be explained by other variables. Commonly, a variable with a variance factor that is higher than four will cause a multicollinearity problem [41–43].

2.1.2. Possible Multicollinearity Issue in Frequency Analysis

Frequency analysis is a statistical method used to predict how often certain values of a variable phenomenon occurs. In hydrology, it is a tool for determining design rainfalls and/or design discharges, which are used as input for designing hydraulic structures. It is also possible to determine the return period of a flood event by analyzing the rainfall or discharge time series data.

The connection between the multicollinearity and the frequency analysis can be found in the so-called rainfall intensity formula. In fact, the rainfall intensity formula is derived for estimating the rainfall intensity under the given return period and rainfall duration. The general form of the rainfall intensity formula is as follows [44]:

$$I = \frac{a + b\log T}{D^n + c} \tag{4}$$

where I is the rainfall intensity (mm/hour), D is the rainfall duration (hour), T is the return period (year), and a, b, c and n are the parameters representing the climate of the study area.

The relation between the return period and rainfall intensity can easily be found in the rainfall intensity formula. First, this equation is solved for T as follows:

$$b\log T = I \times (D^n + c) - a \tag{5}$$

or

$$\log T = \alpha + \beta \times I \tag{6}$$

where α and β are constants. Above equation says that log(T) can be expressed by a linear function of I in case that the rainfall duration D is fixed.

It is also possible to derive a similar equation for the total rainfall depth $R = I \times D$. That is,

$$b\log T = I \times (D^n + c) - a = I \times (DD^{n-1} + c) - a = R \times D^{n-1} + c \times I - a \tag{7}$$

Under the condition that the rainfall intensity I and rainfall duration D are given, the following linear relation can be derived:

$$\log T = \alpha + \beta \times R \tag{8}$$

However, this linear relation assumption is somewhat excessive as the total rainfall depth R is generally proportional to the rainfall duration D. Log(T) may increase rather exponentially as R increases. The relation between the log(T) and D is also similar with that of log(T) and R. As can be found in Equation (4), their relation is not linear.

However, at this moment, the authors assume that their relations are all linear. The authors also realize that this linear assumption is somewhat immoderate or excessive, but we believe this linearization assumption can be used effectively for explaining the possible existence of multicollinearity issue in the multivariate frequency analysis of rainfall events. By adding all these possible relations, the authors assumed that log(T) is expressed as a multiple linear function of I, D and R.

$$\log T = \alpha + \beta_1 \times I + \beta_2 \times D + \beta_3 \times R \tag{9}$$

It is thus obvious that the possible interdependence among three predictors can cause a significant multicollinearity problem even in the frequency analysis. That is, the multicollinearity can cause the same problem as in the regression analysis and hinder the effective estimation of design rainfall or the return period of a given rainfall event based on frequency analysis.

2.2. Copula

In this study, bivariate frequency analysis was conducted by applying copula. Sklar [19] introduced the copula as a tool to derive joint probability distribution (PDF) from several marginal PDFs. Any kind of joint PDF can be derived by applying copula on various marginal PDFs. More information about copula is summarized in [45–47].

This study considered Clayton copula [48], Frank copula [49], Gumbel-Hougaard copula [50], and Gaussian copula [51] for the analysis. Among these copula, Clayton, Frank and Gumbel-Hougaard are classified as Archimedean copula family and Gaussian is in the class of elliptical copula family. These copulas are known for easy application and have many references in hydrology area [7,12,18,20]. Only one parameter, which is referred as θ, is required to derive the joint PDF of copula considered in this study.

Parameter θ for copula can be estimated by calculating Kendall's τ. Kendall's τ is the statistic indicating the rank correlation between two target variables. The equation to calculate Kendall's τ can be expressed as Equation (10).

$$Kendall's\ \tau = \frac{I_n - J_n}{{}_nC_2} \tag{10}$$

In Equation (10), a total number of target data is *n*. If two target data are defined as $(X_i,\ Y_i)$ and $(X_j,\ Y_j)$, I_n is the number of cases that $(X_i - X_j)(Y_i - Y_j) > 0$ and J_n is the number of cases that $(X_i - X_j)(Y_i - Y_j) < 0$. ${}_nC_2$ is the 2-combination from a given set *n*.

Each copula has different relationship between Kendall's τ and parameter θ. For Clayton copula, the relationship Kendall's τ and parameter θ is rather simply defined as $\tau = \frac{\theta}{\theta+2}$. Gumbel-Hougaard copula also has simple relationship as $\tau = 1 - \frac{1}{\theta}$. In the case of Gaussian copula, the relationship is derived as $\tau = \frac{2}{\pi}\arcsin\theta$. The relationship for Frank copula is far complicated. It is defined as follows:

$$\tau = 1 - \frac{4}{\theta} + 4\frac{D_1(\theta)}{\theta} \tag{11}$$

where $D_1(\theta)$ can be calculated by $\int_0^\theta \frac{t}{e^t - 1}dt$.

There is a valid range of parameter θ for each copula and only the parameter θ in the valid range can be applied for the analysis. The parameter θ for the Clayton copula must be defined in $[0,\ \infty)$. The valid range for the Gumbel-Hougaard copula is $[1,\ \infty)$. In the case of Gaussian copula, the parameter θ is valid within $[-1,\ 1]$. Frank copula can be applied with any parameter θ except for the zero.

Equations (12–15) are the joint PDFs of the copula considered in this study. Respectively, Equations (12–15) are for Clayton, Frank, Gumbel-Hougaard and Gaussian. Here, μ_1 and μ_2 are the cumulative probabilities of the bivariate data. These values can be calculated using the marginal PDF. In the case of Φ^{-1}, it is the inverse function of the standard normal distribution.

$$C_{Clayton}(u_1,\ u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta} \tag{12}$$

$$C_{Frank}(u_1,\ u_2) = -\frac{1}{\theta}\ln[1 + \frac{(\exp(-\theta u_1) - 1)(\exp(-\theta u_2) - 1)}{\exp(-\theta) - 1}] \tag{13}$$

$$C_{Gumbel-Houhaard}(u_1,\ u_2) = \exp(-((-\ln(u_1))^\theta + (-\ln(u_2))^\theta\ )^{1/\theta}) \tag{14}$$

$$C_{Gaussian}(u_1,\ u_2) = \int_{-\infty}^{\Phi^{-1}(u_1)}\int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\theta^2}}\exp(-\frac{s^2 + w^2 - 2sw\theta}{2(1-\theta^2)})dsdw \tag{15}$$

Among four copulas, this study selected an optimal one for the analysis. Since there are three pairs of bivariate data, three optimal copulas had to be determined. This study estimated the Akaike information criterion (AIC) to select best copula for analysis. Theoretically, the AIC may not

be applied to test a goodness-of-fit of model, however, this statistic can be used to determine the optimal model among many other admissible models available [18].

Generally, the maximized likelihood statistic is required to calculate the AIC. Alternatively, the mean square error (MSE) between the cumulative probability from copula and that from empirical equation can be used to estimate the AIC [52,53]. This study calculated the MSE to estimate the AIC of each copula. For empirical cumulative probability, the Gringorten plotting position formula [54] was used. Then, the MSE can be calculated by the mean of sum of squared error. Finally, the AIC can be estimated using the following Equation (16).

$$AIC = n\log(MSE) + 2(PAR) \tag{16}$$

In Equation (16), n is the number of data, and PAR is the number of fitted parameter. The AIC calculated by Equation (16) is the penalty statistic. That is, the model with lower AIC can be considered as better model for given data. Therefore, this study determined the optimal copula by selecting the copula with the lowest AIC.

In this study, the optimal copula determined by the lowest AIC was additionally tested with the Kolmogorov-Smirnov (K-S) statistic. The K-S statistic is equal to the maximum absolute difference between the cumulative probability from copula and that from empirical equation [55]. If the calculated K-S statistic is lower than the critical value of the K-S test, the goodness-of-fit of the copula is guaranteed. For the critical value of the K-S test, the significance level of 5% was considered.

Exceedance probability of bivariate data can be calculated with the determined optimal copula. Finally, the return period of bivariate data can then be estimated using the calculated exceedance probability. This study considered three types of return period [12,56]. The type of return period is defined based on the method used to calculate the exceedance probability. First, Equation (17) shows the return period of an OR case. An OR case return period is calculated with the probability that either one or two variables exceed the threshold.

$$T_{or} = \frac{1}{1 - C(F_X(x_i),\ F_Y(y_i))} \tag{17}$$

where $C(F_X(x_i),\ F_Y(y_i))$ is the joint cumulative probability calculated by the copula. Next, an AND case of return period is defined as Equation (18). In this case, the exceedance probability is calculated with the probability that both two variables exceed the threshold.

$$T_{and} = \frac{1}{1 - F_X(x_i) - F_Y(y_i) + C(F_X(x_i),\ F_Y(y_i))} \tag{18}$$

In Equation (18), $F_X(x_i)$ and $F_Y(y_i)$ are the cumulative probabilities calculated by their marginal probability distributions. Finally, Equation (19) is for the COND case of the return period. This return period is based on the event of $X > x$ given $Y = y$.

$$T_{X|Y} = \frac{1}{1 - \frac{\partial}{\partial y} C(F_X(x_i),\ F_Y(y_i))} \tag{19}$$

In this study, the bivariate frequency analysis was conducted to calculate an AND case return period. In addition, the return period for OR and COND cases were considered to validate the results of the bivariate frequency analysis.

3. Annual Maximum Rainfall Events in Seoul, Korea

In this study, the annual maximum rainfall events of Seoul from 1961 to 2010 were secured for bivariate frequency analysis [57]. Freund's bivariate exponential mixture distribution was applied to select the annual maximum rainfall events. In this case, the total rainfall depth and mean rainfall intensity were used for frequency analysis. The parameters of Freund's bivariate exponential

distribution were estimated annually. It is generally known that the annually estimated parameters are better to consider the different rainfall characteristics of wet and dry years [58]. Since the number of independent rainfall events is more than 30 every year, there was no serious problem in estimating the parameters annually. Based on frequency analysis result, annual maximum rainfall event with the highest return period for each year was determined.

The basic statistics of the rainfall duration, total rainfall depth, and mean rainfall intensity of the rainfall events are summarized in Table 1. The mean of the rainfall duration was 21.7 hour and that of the total rainfall depth was 172.1 mm. For the mean rainfall intensity, the mean was calculated to be 12.3 mm/hour. In the case of the standard deviations, they were calculated to be 19.0 hour, 102.9 mm, 7.7 mm/hour for the rainfall duration, total rainfall depth, and mean rainfall intensity, respectively. The ranges of the rainfall duration, total rainfall depth, and mean rainfall intensity were 2 to 94 hour, 39.4 mm to 446.0 mm and 2.9 mm/hour to 32.5 mm/hour, respectively.

Table 1. The means, standard deviations, and ranges of three components of the annual maximum rainfall events observed in Seoul, Korea.

	Mean	Standard deviation	Range
Rainfall duration (hour)	21.7	19.0	2.0–94.0
Total rainfall depth (mm)	172.1	102.9	39.4–446.0
Mean rainfall intensity (mm/hour)	12.3	7.7	2.9–32.5

Before conducting statistical analysis including frequency analysis, it is essential to check the homogeneity and stationarity of the data. In this study, the Bartlett's test has been conducted to investigate the homogeneity of the components of independent rainfall events. Further, the authors found that the null hypothesis of homogeneity was rejected with p-value less than 0.001. For this reason, t-test in this study was conducted under unequal variances. Moreover, the Box-Pierce and Ljung-Box test were applied to examine the null hypothesis of independence among rainfall events. The p-values for both Box-Pierce and Ljung-Box test were all estimated to be higher than 0.05 at lags up to 10. Based on this result, the authors could assume the independence of the rainfall events considered in this study. Finally, in the case of stationarity, the Augmented Dickey-Fuller test was applied on each component of the independent rainfall events. As a result, all of three components were figured out to be stationary with the p-value less than 0.05. The bivariate frequency analysis in this study could thus be proceeded under assumption of stationarity.

The return period of each annual maximum rainfall event was calculated using the rainfall intensity formula (RIF) suggested by Bernard [59].

$$I = \frac{KT^x}{D^b} \tag{20}$$

where I is the rainfall intensity (mm/hour), D is the rainfall duration (hour), T is the return period (year), and K, x and b are the parameters representing the climate of the study area. The parameters K, x and b were determined as 8.0, 0.5 and 0.5, respectively, to agree with the statistical characteristics of the observed annual maximum rainfall events observed in Seoul, Korea [57]. With these parameters, the return period T for each annual maximum rainfall event could be calculated by solving Equation (20).

Figure 2 shows the basic components of annual maximum rainfall events and the calculated return periods. The longest return period was calculated to be 129.7 years for the rainfall event observed in 1972, whose duration was 24.0 hours and total rainfall depth 446.0 mm (i.e., mean rainfall intensity 18.6 mm/hour). Among 50 annual maximum rainfall events considered in this

study, 4 events were found to have return periods less than 10 years (i.e., $T < 10$), 30 events less than 30 years (i.e., $10 < T < 30$), 8 events less than 50 years (i.e., $30 < T < 50$), and the remaining 8 events longer than 50 years (i.e., $T > 50$).

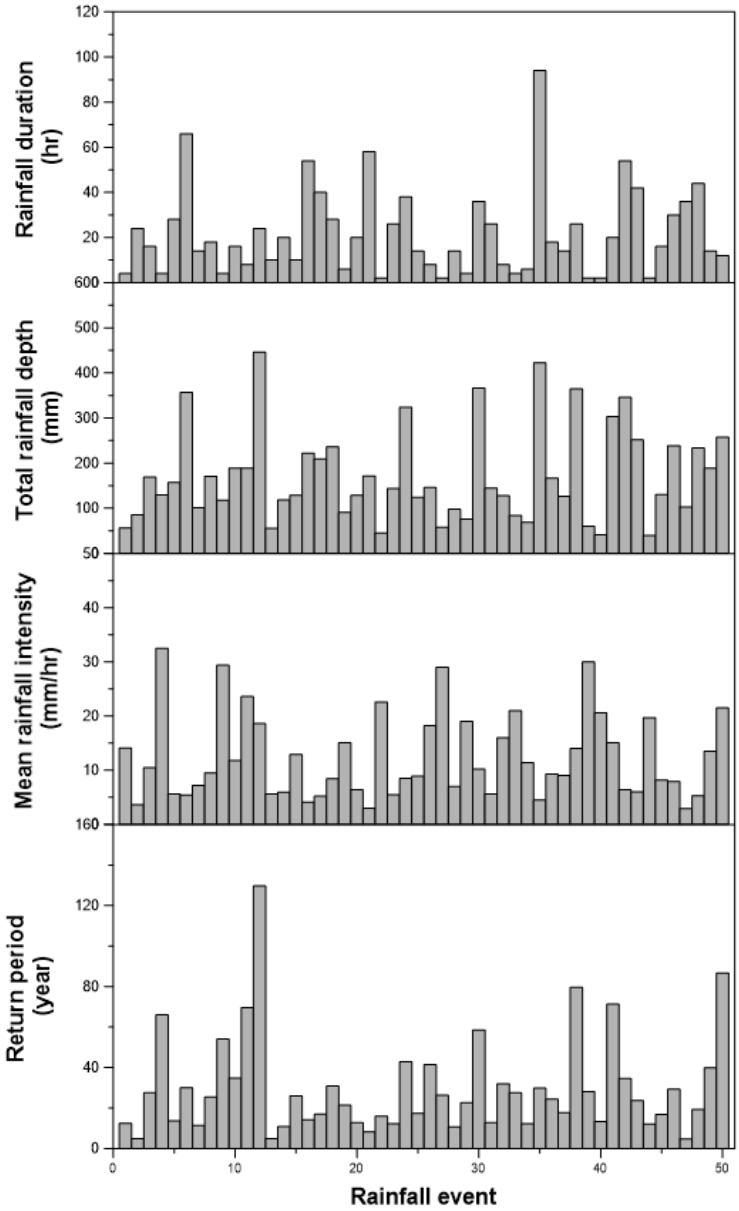


Figure 2. Variation of the basic components of the annual rainfall events observed in Seoul Korea and their return periods.

4. Evaluation of Multicollinearity Problem with Observed Data

4.1. Results of Bivariate Frequency Analysis

First, the Kendall's τ was calculated to estimate the parameter of the copula. The Kendall's τ for the total rainfall depth and rainfall duration was 0.57 and that for the mean rainfall intensity and the rainfall duration was −0.61. In the case of the total rainfall depth and the mean rainfall intensity, Kendall's τ was −0.18. As expected, a positive correlation was estimated between the total rainfall depth and the rainfall duration and a negative correlation was estimated between the mean rainfall intensity and the rainfall duration. These three Kendall's τ in this study were also found to be statistically significant based on the *p*-value of the t-test. The *p*-value for total rainfall depth and rainfall duration was estimated to be lower than 0.001. It was 0.0020 for the mean rainfall intensity

and the rainfall duration, and it was also lower than 0.001 for the total rainfall depth and the mean rainfall intensity.

Table 2 shows the parameter θ estimated using the relationship between the parameter θ and the Kendall's τ. In this table, NA represents the invalid estimate of the parameter θ because it is beyond the valid range. Especially, the valid range of the parameter θ for the Clayton and Gumbel-Hougaard models is $\theta \geq 1$.

Table 2. The parameter θ estimated for each data pair with different copulas.

Copula model	Total rainfall depth and rainfall duration	Mean rainfall intensity and rainfall duration	Total rainfall depth and mean rainfall intensity
Clayton	2.65	NA	NA
Frank	7.17	−1.19	−1.02
Gumbel-Hougaard	2.32	NA	NA
Gaussian	0.78	−0.82	−0.28

This study determined the PDF for each variable of the annual maximum rainfall event to conduct the BFA. As a result, the exponential distribution, the Gumbel distribution and the gamma distribution were chosen for the rainfall duration, total rainfall depth and mean rainfall intensity, respectively; these are frequently used in the modeling of rainfall events [60–62]. Each probability distribution was goodness-of-fit tested by the Kolmogorov-Smironov test computed via a Monte-Carlo procedure. The *p*-value for the rainfall duration with exponential distribution was calculated as 0.5216, and that for the total rainfall depth with Gumbel distribution was 0.7800. Finally, the *p*-value for the mean rainfall intensity with gamma distribution was 0.7023. The high *p*-values for these three probability distributions indicate that they cannot be rejected at 0.05 significance level.

Before the optimal copula is determined, this study conducted the goodness-of-fit test using the function of "gofCopula()" provided by the R package "copula"[63]. The *p*-values for the three cases of bivariate analysis could be calculated as Table 3. As can be seen in this table, the *p*-values of all copulas in this study were high enough to confirm the goodness-of-fit. That is, all copulas were found to be statistically significant.

Table 3. P-values calculated for three bivariate data pairs of the annual maximum rainfall events observed in Seoul, Korea.

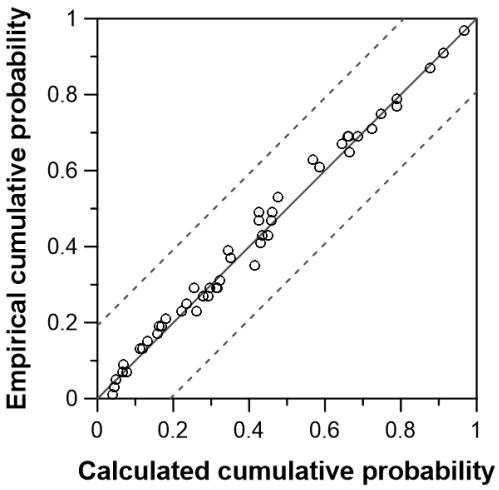
Case	Clayton	Frank	Gumbel-Hougaard	Gaussian
Total rainfall depth and rainfall duration	0.4461	0.6009	0.5120	0.9306
Mean rainfall intensity and rainfall duration	NA	0.3841	NA	0.6199
Total rainfall depth and mean rainfall intensity	NA	0.3012	NA	0.3412

Next, this study estimated MSE and the AIC to select optimal copula for each bivariate data pair. Estimated MSE and AIC are summarized in Table 4. The optimal copula was determined as the copula with the lowest MSE and AIC. Therefore, the optimal copula for the total rainfall depth and the rainfall duration was found to be Clayton copula. In the case of the mean rainfall intensity and the rainfall duration, the Gaussian copula was selected as the optimal copula. For the total rainfall depth and the mean rainfall intensity, the Frank copula was chosen.

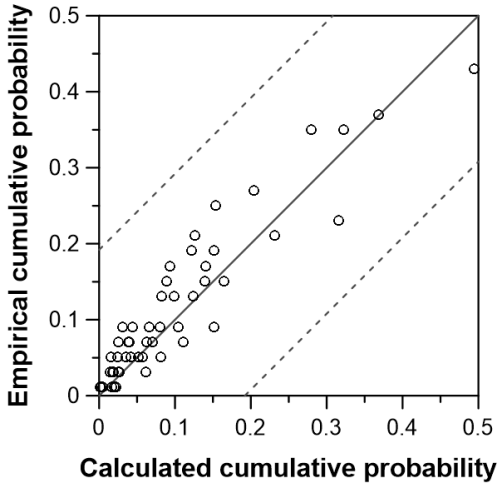
Table 4. The mean square errors (MSEs) and Akaike information criterions (AICs) calculated for three bivariate data pairs of the annual maximum rainfall events observed in Seoul, Korea.

	Case	Clayton	Frank	Gumbel-Hougaard	Gaussian
MSE	Total rainfall depth and rainfall duration	0.000352	0.000916	0.0216	0.000881
	Mean rainfall intensity and rainfall duration	NA	0.00449	NA	0.00161
	Total rainfall depth and mean rainfall intensity	NA	0.00132	NA	0.00168
AIC	Total rainfall depth and rainfall duration	−156.206	−149.911	−81.323	−150.757
	Mean rainfall intensity and rainfall duration	NA	−115.382	NA	−137.620
	Total rainfall depth and mean rainfall intensity	NA	−141.971	NA	−136.681

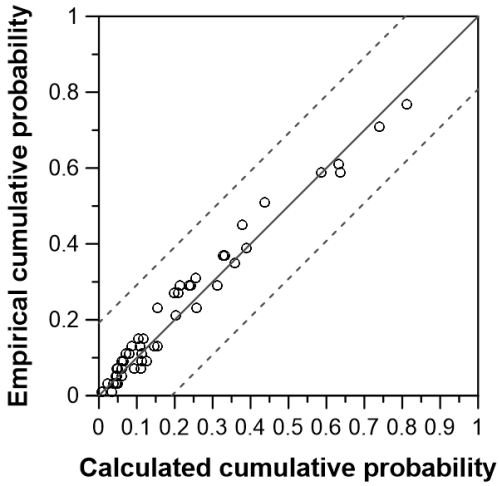
Figure 3 is the scatter plot indicating the cumulative probability calculated by the Gringorten plotting position formula and the optimal copula. In Figure 3, dotted lines represent the critical value to examine the goodness-of-fit of copula. In this study, the critical value was calculated to be 0.192 for n = 50 and 5% of the significance level. All the points from three bivariate data pairs found to be within the upper and lower limit of K-S test. Therefore, all the optimal copula in this study can be regarded as verified one under K-S test.



(**a**)



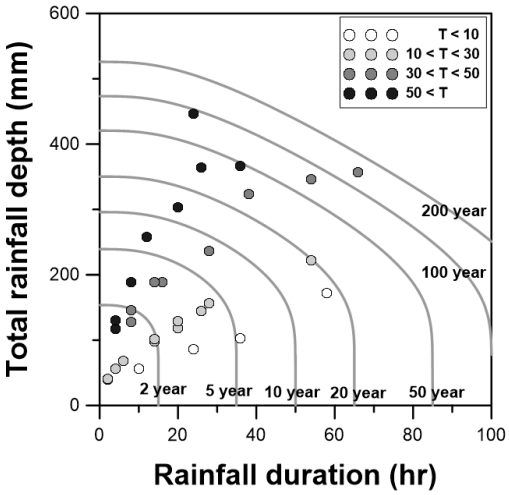
(**b**)



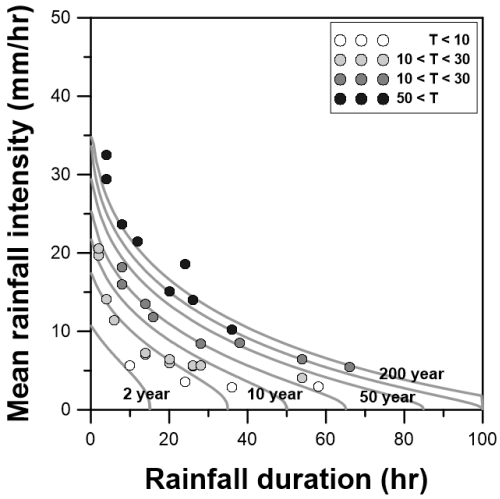
(**c**)

Figure 3. K-plot and Kolmogorov-Smirnov (K-S) test criteria of three optimal copula models for the annual maximum rainfall events observed in Seoul, Korea. (**a**) Clayton copula model; (**b**) Gaussian copula model; (**c**) Frank copula model.

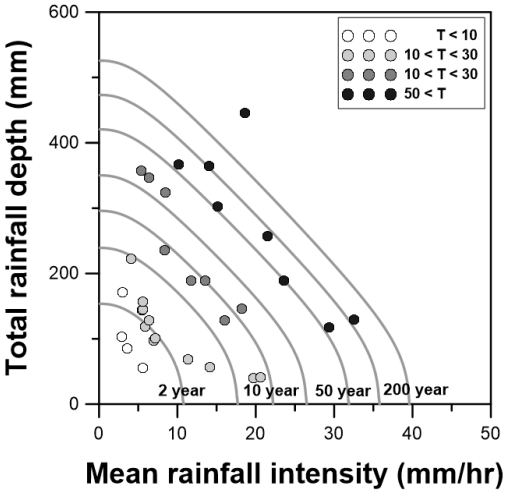
The BFA was then conducted with the optimal copula model that was selected for each bivariate data of the annual maximum rainfall events. As a result, the AND return period T_{and} was estimated for each rainfall event, which was then compared with the return period calculated by the RIF (Equation (20)). In Figure 4, the return periods estimated using the BFA are presented by contour lines and those calculated using the RIF by solid circles. In addition, for easier comparison, different darkness was applied to fill the circle. For example, white color was applied to those rainfall events whose return periods were calculated to be less than 10 years by the RIF (i.e., $T < 10$). The light grey was applied to those rainfall events whose return periods were less than 30 years (i.e., $10 < T < 30$) and the dark grey was applied to those rainfall events whose return periods were less than 50 years (i.e., $30 < T < 50$). Finally, black color was applied to those rainfall event whose return period was calculated to be longer than 50 years (i.e., $T > 50$).



(**a**)



(**b**)



(**c**)

Figure 4. Comparison of return periods estimated by the bivariate frequency analysis (BFA) and those calculated by the rainfall intensity formula (RIF). (**a**) Total rainfall depth and rainfall duration; (**b**) Mean rainfall intensity and rainfall duration; (**c**) Total rainfall depth and mean rainfall intensity.

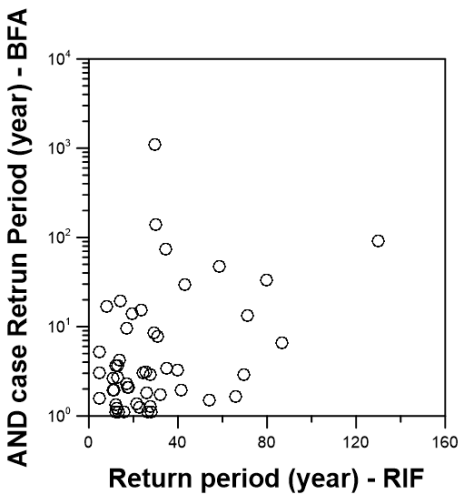
Figure 4 shows the way that the return period of a rainfall event can be very different from that which has been estimated from the BFA. As the definitions of the two return periods in this figure are not identical, the results of frequency analysis can also be different [30,64]. Still, the authors expected to find a certain extent of consistency between the return period calculated by the RIF and the return period estimated by the BFA. However, this expectation was not fulfilled in the case where the total rainfall depth and the rainfall duration were considered as the two variables, as can be seen Figure 4a. The return periods of the rainfall events with the same return period calculated by the RIF could be totally different. That is, there are so many cases where the return period estimated by BFA is totally different from the return period calculated by the RIF. This finding indicates that the result of the BFA for the total rainfall depth and the rainfall duration does not reflect the return period calculated by the RIF.

The return period calculated by the BFA in Figure 4b,c is somewhat proportional to the return period calculated by the RIF. This is more vivid in Figure 4c for which the total rainfall depth and the mean rainfall intensity were considered. This trend is still valid, even though it is not strong, for the case where the mean rainfall intensity and the rainfall duration were considered (Figure 4b). Obviously, when the rainfall duration is involved, the proportional trend between the return period calculated by the RIF and the return period estimated by the BFA becomes weak.

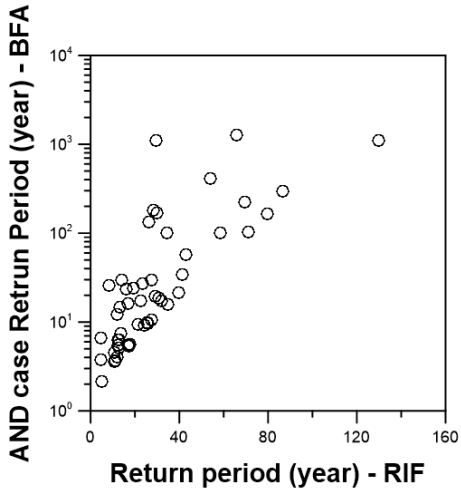
4.2. Effect of Multicollinearity on the Estimated Return Periods

The authors assumed that the results of the frequency analysis (i.e., the return period) for the same rainfall event should be the same regardless of the selected two variables (out of three). For example, this study considered a rainfall event with its duration of 2 hours and rainfall intensity of 50 mm/hour (i.e., the total rainfall depth is 100 mm in 2 hours). Among three variables, the rainfall intensity, rainfall duration and total rainfall depth, one is redundant as it is calculated by the other two. Now, we can select any two variables for the bivariate frequency analysis. Regardless of the two variables selected, it is expected to obtain a similar result of bivariate frequency analysis (i.e., the return period) for the given rainfall event.

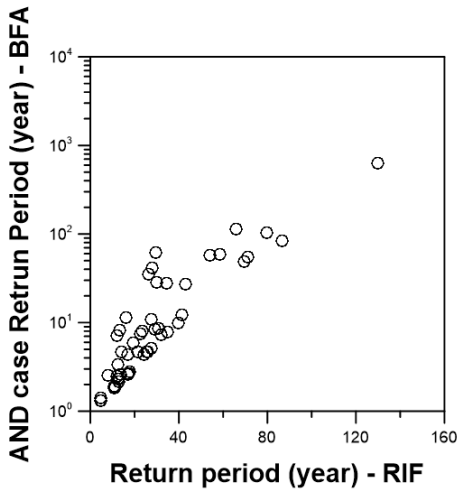
To confirm this assumption, scatter plots were made to compare the return period calculated by RIF and the return period estimated by the BFA (Figure 5). In this figure, the *x*-axis represents the return period calculated by the RIF on a linear scale and the *y*-axis represents the resulting return period estimated by the BFA on a logarithmic scale. The authors tested all the linear, semi-log and log-log graphs and found that the semi-log was the best to search a possible relation between the two variables.



(**a**)



(**b**)



(**c**)

Figure 5. Scatter plots of return periods estimated by the BFA and those calculated by the RIF. (**a**) Total rainfall depth and rainfall duration; (**b**) Mean rainfall intensity and rainfall duration; (**c**) Total rainfall depth and mean rainfall intensity.

When the three scatter plots were compared, firstly it was found that the three results of the BFA are not consistent at all. The overall result is very dependent on the two variables that are considered in the frequency analysis. Even though the two variables were selected from the same rainfall event data, the results of the frequency analysis are totally different from each other. A closer look at this figure shows that, as can be seen in Figure 5a, the return periods that were estimated from the analysis of the total rainfall depth and the rainfall duration do not show any obvious increasing trends with respect to the return period. That is, the return period calculated by the RIF did not affect the return period. A rather consistent increasing trend was, however, found in Figure 5b,c. Between the cases that are shown in Figure 5b,c, Figure 5c more effectively shows the way that the return period estimated by the BFA is more similar with the return period calculated by the RIF and that the range of the return period estimated by the BFA is also much smaller than that of Figure 5b.

The above findings are also summarized in Table 5 for a comparison. As shown in Table 5, the means of the three cases are very different from one another. The mean for the case where the total rainfall depth and the mean rainfall intensity were considered is the lowest at approximately 31.1

years, but it is rather high at approximately 34.1 years for the case for which the total rainfall depth and the rainfall duration were considered. The result is extreme for the case for which the mean rainfall intensity and the rainfall duration were considered, which shows that the mean return period is higher than 100 years. Even in the case for which the total rainfall depth and the mean rainfall intensity were considered, the range of the estimated return periods is rather absurd.

Table 5. The mean and range of the return period (years) for annual maximum rainfall events observed in Seoul, Korea.

Case	Mean	Range
Total rainfall depth and rainfall duration	34.1	1.1–1,105.0
Mean rainfall intensity and rainfall duration	118.2	2.1–1,289.4
Total rainfall depth and mean rainfall intensity	31.1	1.3–629.6

In fact, above results could be expected by comparing the variance inflation factors estimated for the three components that are considered in this study. The variance inflation factor of the rainfall duration was estimated as 4.87 and that of the total rainfall depth was estimated as 3.65. In the case of the mean rainfall intensity, the variance inflation factor is 1.81. As a result, the rainfall duration is highly vulnerable to the multicollinearity problem because its variance inflation factor is higher than four. It could also be concluded that the BFA for which the total rainfall depth and the mean rainfall intensity are used is more reasonable than the use of the bivariate data including the rainfall duration. In case of considering the rainfall duration, high uncertainty about the estimation of return period cannot be avoided.

However, a note of caution is due here. In fact, several ties (i.e., repeated values) are observed for the rainfall duration. As discussed in [65], on the one hand, an issue of model identifiability may arise (both considering the marginals and the copula at play), and on the other hand, the assessment of the return periods may be affected. Several randomization techniques have been outlined in [66,67] in order to deal with ties, but in the present work we pass over the issue.

The explanation in the previous paragraph could also be confirmed by the regression analysis based on Equation (9). First, the regression equation derived was as follows:

$$\log T = -0.74 + 0.0031D + 0.0045R + 0.068I \tag{21}$$

The significance of the above equation was found to be very high (p-value $< 2\times10^{-16}$). However, the significance of those predictors varied. The p-values for I and R were found to be much smaller than that of D. This result indicates that there is some problem in relating the return period with all those three predictors.

The authors thus considered the multicollinearity issue to decrease the number of predictors. One variable, the rainfall duration, was excluded in this step as its variance inflation factor was very high to be 4.87. In general, the variable with its variance inflation factor higher than 4 is excluded in the multivariate regression analysis [43–45]. Now the regression equation with two predictors, the total rainfall depth R and the mean rainfall intensity I, was derived again.

$$\log T = -0.68 + 0.0049R + 0.064I \tag{22}$$

The significance for above Equation (22) was still very high (p-value $< 2\times10^{-16}$). Furthermore, this equation did not contain any variable statistically insignificant. This result indicates that the multicollinearity issue was removed. Furthermore, these results indicate that it is important in the multivariate frequency analysis of rainfall events to consider the multicollinearity issue. This multicollinearity among predictors can hinder the effective estimation of the return period of a given rainfall event.

Additionally, the choice of the method to calculate return period has a significant influence on the outcome [68]. For this reason, the OR case return period and the COND case return period were also estimated using total rainfall depth and mean rainfall intensity.

Figure 6 shows the estimated OR case return period and conditional return period with the return period calculated by the RIF. In the case of OR case return period, there were many results overlapped by other results. Still, there was a consistency between the OR case return periods calculated by BFA and those calculated by the RIF. There was also a similar tendency of results for the case of conditional return period. Therefore, the results of OR case return period and conditional return period found to support the result of this study derived by considering multicollinearity.

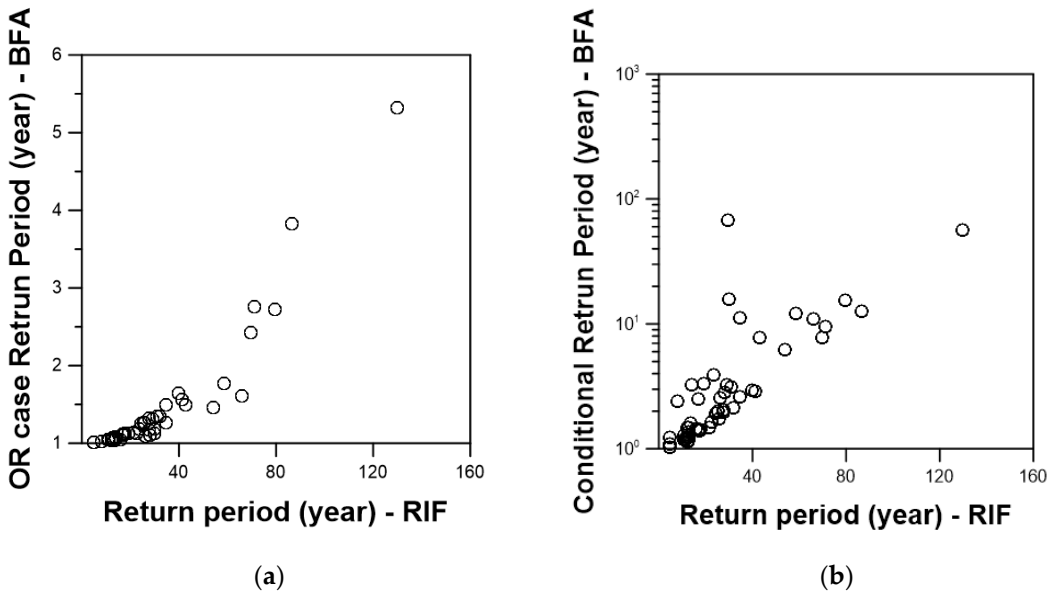


Figure 6. Scatter plots of OR case and Conditional return periods estimated by the BFA and those calculated by the RIF. (**a**) OR case return period; (**b**) Conditional return period.

In the study from Kroll and Song [69], the regression model results with/without multicollinearity issue have been compared based on the prediction ability. The variables with the high variance inflation factor have been screened for the regression model. As a result. it was concluded that the negative effect of multicollinearity could be magnified without screening, as the size of sample (or degree of freedom) decreased. In this study, the authors suggest selecting the variables having small variation inflation factor for the bivariate frequency analysis. Actually, this conclusion about excluding variables with the multicollinearity problem well agrees with the conclusion from Kroll and Song [69].

5. Conclusions

A rainfall event simplified by a rectangular pulse is defined by three components: the rainfall duration, the total rainfall depth, and mean rainfall intensity. However, as the mean rainfall intensity can be calculated by the total rainfall depth divided by the rainfall duration, any two components can fully define the rainfall event (i.e., one component must be redundant). In this study, the selection issue of the two variables for the BFA was discussed. This issue is possibly associated with the multicollinearity in multivariate regression analysis. That is, high interdependency among predictors cause the increase of the variance of the regression coefficient and finally lead to the low quality of resulting estimate. To verify this assumption, the annual maximum rainfall events collected in Seoul, Korea was analyzed as an example. The BFA was repeated three times with different pairs of the two variables among the three components of the rainfall events, i.e., rainfall duration, total rainfall depth, and mean rainfall intensity. Finally, the return periods of the rainfall events were estimated and compared with those calculated by the RIF.

As a result of the BFA of the annual maximum rainfall events, the return period of a rainfall event was estimated very differently for each event among the three cases that are considered in this study. For example, return period of the rainfall event observed in 2010 range from 6.6 years to 295.0 years depending on the two variables selected. This dependency of the resulted return period on the selected variable pair could be explained effectively using the multicollinearity among the variables. Among the three variables that are considered in this study, the rainfall duration caused the most serious multicollinearity problem. As a result, the resulting return period of the BFA could be most biased when the rainfall duration was considered in the BFA. This result indicates that the result of the BFA is most realistic when the total rainfall depth and the mean rainfall intensity were considered. The return periods estimated by the BFA are very consistent with return periods calculated by the RIF.

In conclusion, the result of the BFA with the total rainfall depth and the mean rainfall intensity was confirmed as the most reasonable case among the three cases for which different pairs of the two variables were considered. This result also completely agreed with the multicollinearity issue among the correlated predictors. Similar results with this study are expected in other regions, unless the correlation structure among mean rainfall intensity, rainfall duration and the total rainfall depth is totally different. However, it is also true that the result can vary a bit region by region, which is obviously dependent upon the regional climate.

Author Contributions: C.Y. had done conceptualization and writing review and editing E.C. had done data curation and visualization and writing original draft preparation.

Funding: This work is supported by the Korea Environmental Industry & Technology Institute (KEITI) grant funded by the Ministry of Environment (Grant RE201902084) and a grant (19CTAP-C143641-02) from Infrastructure and transportation technology promotion research Program funded by Ministry of Land Infrastructure and Transport of Korean government.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. De Haan, L.; De Ronde, J. Sea and wind: Multivariate extremes at work. *Extremes* **1998**, *1*, 7, doi:10.1023/A:1009909800311.
2. Yue, S. Applicability of the Nagao–Kadoya bivariate exponential distribution for modeling two correlated exponentially distributed variates. *Stoch. Environ. Res. Risk Assess.* **2001**, *15*, 244–260, doi:10.1007/s004770100069.
3. Yue, S.; Wang, C. A comparison of two bivariate extreme value distributions. *Stoch. Environ. Res. Risk Assess.* **2004**, *18*, 61–66, doi:10.1007/s00477-003-0124-x.
4. Yue, S. Joint probability distribution of annual maximum storm peaks and amounts as represented by daily rainfalls. *Hydrol. Sci. J.* **2000**, *45*, 315–326, doi:10.1080/02626660009492327.
5. Herr, H.D.; Krzysztofowicz, R. Generic probability distribution of rainfall in space: The bivariate model. *J. Hydrol.* **2005**, *306*, 234–263, doi:10.1016/j.jhydrol.2004.09.011.
6. Nadarajah, S. A bivariate pareto model for drought. *Stoch. Environ. Res. Risk Assess.* **2009**, *23*, 811–822, doi:10.1007/s00477-008-0260-4.
7. Mirabbasi, R.; Fakheri-Fard, A.; Dinpashoh, Y. Bivariate drought frequency analysis using the copula method. *Theor. Appl. Climatol.* **2012**, *108*, 191–206, doi:10.1007/s00704-011-0524-7.
8. Yoo, J.; Kwon, H.-H.; Kim, T.-W.; Ahn, J.-H. Drought frequency analysis using cluster analysis and bivariate probability distribution. *J. Hydrol.* **2012**, *420*, 102–111, doi:10.1016/j.jhydrol.2011.11.046.
9. Zhang, L.; Singh, V. Bivariate flood frequency analysis using the copula method. *J. Hydrol. Eng.* **2006**, *11*, 150–164, doi:10.1061/(ASCE)1084-0699(2006)11:2(150).
10. Grimaldi, S.; Serinaldi, F. Asymmetric copula in multivariate flood frequency analysis. *Adv. Water Resour.* **2006**, *29*, 1155–1167, doi:10.1016/j.advwatres.2005.09.005.
11. Chebana, F.; Ouarda, T.B. Depth-based multivariate descriptive statistics with hydrological applications. *J. Geophys. Res. Atmos.* **2011**, *116*, D10120.
12. De Michele, C.; Salvadori, G.; Canossi, M.; Petaccia, A.; Rosso, R. Bivariate statistical approach to check adequacy of dam spillway. *J. Hydrol. Eng.* **2005**, *10*, 50–57, doi:10.1061/(ASCE)1084-0699(2005)10:1(50).

13. Park, C.-S.; Yoo, C.-S.; Jun, C.-H. Bivariate Rainfall Frequency Analysis and Rainfall-runoff Analysis for Independent Rainfall Events. *J. Korea Water Resour. Assoc.* **2012**, *45*, 713–727, doi:10.3741/JKWRA.2012.45.7.713.
14. Yoon, P.; Kim, T.-W.; Yoo, C. Rainfall frequency analysis using a mixed GEV distribution: A case study for annual maximum rainfalls in South Korea. *Stoch. Environ. Res. Risk Assess.* **2013**, *27*, 1143–1153, doi:10.1007/s00477-012-0650-5.
15. De Michele, C.; Salvadori, G. A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas. *J. Geophys. Res.* **2003**, *108*, doi:10.1029/2002JD002534.
16. Bacchi, B.; Becciu, G.; Kottegoda, N.T. Bivariate exponential model applied to intensities and durations of extreme rainfall. *J. Hydrol.* **1994**, *155*, 225–236, doi:10.1016/0022-1694(94)90166-X.
17. Yue, S. A bivariate gamma distribution for use in multivariate flood frequency analysis. *Hydrol. Process.* **2001**, *15*, 1033–1045, doi:10.1002/hyp.259.
18. Zhang, L.; Singh, V.P. Bivariate rainfall frequency distributions using Archimedean copulas. *J. Hydrol.* **2007**, *332*, 93–109, doi:10.1016/j.jhydrol.2006.06.033.
19. Sklar, M. Fonctions de repartition an dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris* **1959**, *8*, 229–231.
20. Fu, G.; Butler, D. Copula-based frequency analysis of overflow and flooding in urban drainage systems. *J. Hydrol.* **2014**, *510*, 49–58, doi:10.1016/j.jhydrol.2013.12.006.
21. Lee, D.R.; Jeong, S.M. A SpatialTemporal Characteristics of Rainfall in the Han River Basin. *J. Korean Assoc. Hydrol. Sci.* **1992**, *25*, 75–85.
22. Yoo, C.; Park, C. Comparison of Annual Maximum Rainfall Series and Annual Maximum Independent Rainfall Event Series. *J. Korea Water Resour. Assoc.* **2012**, *45*, 431–444.
23. Rodriguez-Iturbe, I.; De Power, B.F.; Valdes, J. Rectangular pulses point process models for rainfall: Analysis of empirical data. *J. Geophys. Res. Atmos.* **1987**, *92*, 9645–9656.
24. Onof, C.; Wheater, H.S. Modelling of British rainfall using a random parameter Bartlett-Lewis rectangular pulse model. *J. Hydrol.* **1993**, *149*, 67–95, doi:10.1016/0022-1694(93)90100-N.
25. Smithers, J.; Pegram, G.; Schulze, R. Design rainfall estimation in South Africa using Bartlett–Lewis rectangular pulse rainfall models. *J. Hydrol.* **2002**, *258*, 83–99, doi:10.1016/S0022-1694(01)00571-6.
26. Yoo, C.; Kim, D.; Kim, T.-W.; Hwang, K.-N. Quantification of drought using a rectangular pulses Poisson process model. *J. Hydrol.* **2008**, *355*, 34–48, doi:10.1016/j.jhydrol.2008.02.025.
27. Burton, A.; Fowler, H.; Blenkinsop, S.; Kilsby, C. Downscaling transient climate change using a Neyman–Scott Rectangular Pulses stochastic rainfall model. *J. Hydrol.* **2010**, *381*, 18–32, doi:10.1016/j.jhydrol.2009.10.031.
28. Gräler, B.; van den Berg, M.; Vandenberghe, S.; Petroselli, A.; Grimaldi, S.; De Baets, B.; Verhoest, N. Multivariate return periods in hydrology: A critical and practical review focusing on synthetic design hydrograph estimation. *Hydrol. Earth Syst. Sci.* **2013**, *17*, 1281–1296, doi:10.5194/hess-17-1281-2013.
29. Jain, A.; Vinnarasi, R.; Dhanya, C. Intensity-Duration-Frequency Relationship using Event Based Approach. *J. Agroecol. Nat. Resour. Manag.* **2015**, *2*, 178–180.
30. Chong, I.G.; Jun, C.-H. Performance of some variable selection methods when multicollinearity is present. *Chemom. Intell. Lab. Syst.* **2005**, *78*, 103–112, doi:10.1016/j.chemolab.2004.12.011.
31. Silvey, S. Multicollinearity and imprecise estimation. *J. R. Stat. Soc. Ser. B Methodol.* **1969**, *31*, 539–552.
32. Farrar, D.E.; Glauber, R.R. Multicollinearity in regression analysis: The problem revisited. *Rev. Econ. Stat.* **1967**, *49*, 92–107.
33. Cortina, J.M. Interaction, nonlinearity, and multicollinearity: Implications for multiple regression. *J. Manag.* **1993**, *19*, 915–922, doi:10.1016/0149-2063(93)90035-L.
34. Wheeler, D.; Tiefelsdorf, M. Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *J. Geogr. Syst.* **2005**, *7*, 161–187, doi:10.1007/s10109-005-0155-6.
35. Harvey, A. Some comments on multicollinearity in regression. *Appl. Stat.* **1977**, *26*, 188–191, doi:10.2307/2347027.
36. Morrow-Howell, N. The M word: Multicollinearity in multiple regression. *Soc. Work Res.* **1994**, *18*, 247–251, doi:10.1093/swr/18.4.247.
37. Viswesvaran, C. Multiple Regression in Behavioral Research: Explanation and Prediction. *Pers. Psychol.* **1998**, *51*, 223.
38. Allison, P.D. *Multiple Regression: A Primer*; Pine Forge Press: Newbury Park, CA, USA, 1999.

39. Kutner, M.; Nachtsheim, C.; Neter, J.J. *Applied Linear Regression Models*; McGraw-Hill: New York, NY, USA, 2004.
40. Robinson, C.; Schumacker, R.E. Interaction effects: Centering, variance inflation factor, and interpretation issues. *Mult. Linear Regres. Viewp.* **2009**, *35*, 6–11.
41. Demuth, S.; Steffensmeier, D. Ethnicity effects on sentence outcomes in large urban courts: Comparisons among White, Black, and Hispanic defendants. *Soc. Sci. Q.* **2004**, *85*, 994–1011, doi:10.1111/j.0038-4941.2004.00255.x.
42. O'brien, R.M. A caution regarding rules of thumb for variance inflation factors. *Qual. Quant.* **2007**, *41*, 673–690, doi:10.1007/s11135-006-9018-6.
43. Gilliam, J.; Chatterjee, S.; Grable, J. Measuring the perception of financial risk tolerance: A tale of two measures. *J. Financ. Counsel. Plan*. **2010**, *21*, 1–14.
44. Lee, W.H.; Park, S.D.; Choi, S.Y. A derivation of the typical probable rainfall intensity formula in Korea. *J. Korean Soc. Civ. Eng.* **1993**, *13*, 115–120.
45. Nelsen, R.B. *An Introduction to Copulas*; Springer Science & Business Media: Berlin, Germany, 2007; doi:10.1007/0-387-28678-0.
46. Genest, C.; Favre, A.-C. Everything you always wanted to know about copula modeling but were afraid to ask. *J. Hydrol. Eng.* **2007**, *12*, 347–368, doi:10.1061/(ASCE)1084-0699(2007)12:4(347).
47. Salvadori, G.; De Michele, C.; Kottegoda, N.T.; Rosso, R. *Extremes in Nature: An Approach Using Copulas*; Springer Science & Business Media: Berlin, Germany, 2007; doi:10.1007/1-4020-4415-1.
48. Clayton, D.G. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **1978**, *65*, 141–151, doi:10.1093/biomet/65.1.141.
49. Frank, M.J. On the simultaneous associativity of F (x, y) and x+y− F (x, y). *Aequ. Math.* **1979**, *19*, 194–226.
50. Hutchinson, T.P.; Lai, C.D. *Continuous Bivariate Distributions Emphasising Applications*; Rumsby Scientific Publishing: Adelaide, Australia, 1990.
51. Schmidt, R.; Stadtmüller, U. Non-parametric estimation of tail dependence. *Scand. J. Stat.* **2006**, *33*, 307–335, doi:10.1111/j.1467-9469.2005.00483.x.
52. Brooks, C.; Burke, S.P. Information criteria for GARCH model selection. *Eur. J. Financ.* **2003**, *9*, 557–580, doi:10.1080/1351847021000029188.
53. Juneja, V.K.; Melendres, M.V.; Huang, L.; Gumudavelli, V.; Subbiah, J.; Thippareddi, H. Modeling the effect of temperature on growth of Salmonella in chicken. *Food Microbiol.* **2007**, *24*, 328–335, doi:10.1016/j.fm.2006.08.004.46.
54. Gringorten, I.I. A plotting rule for extreme probability paper. *J. Geophys. Res.* **1963**, *68*, 813–814, doi:10.1029/JZ068i003p00813.
55. Yevjevich, V. *Probability and Statistics in Hydrology*; Water Resources Publications, LLC: Highlands Ranch, CO, USA, 1972.
56. Salvadori, G.; De Michele, C. On the use of copulas in hydrology: Theory and practice. *J. Hydrol. Eng.* **2007**, *12*, 369–380, doi:10.1061/(ASCE)1084-0699(2007)12:4(369).
57. Yoo, C.; Park, M.; Kim, H.J.; Jun, C. Comparison of annual maximum rainfall events of modern rain gauge data (1961–2010) and Chukwooki data (1777–1910) in Seoul, Korea. *J. Water Clim. Chang.* **2018**, *9*, 58–73, doi:10.2166/wcc.2017.110.
58. Park, M.; Yoo, C.; Kim, H.; Jun, C. Bivariate frequency analysis of annual maximum rainfall event series in Seoul, Korea. *J. Hydrol. Eng.* **2013**, *19*, 1080–1088, doi:10.1061/(ASCE)HE.1943-5584.0000891.
59. Bernard, M.M. Formulas for rainfall intensities of long duration. *Trans. Am. Soc. Civ. Eng.* **1932**, *96*, 592–606.
60. Brutsaert, W. *Hydrology: An Introduction*; Cambridge University Press: Cambridge, UK, 2005.
61. Gupta, S.K. *Modern Hydrology and Sustainable Water Development*; John Wiley & Sons: Hoboken, NJ, USA, 2011.
62. Naghettini, M. *Fundamentals of Statistical Hydrology*; Springer: Berlin, Germany, 2017, doi:10.1007/978-3-319-43561-9.53.
63. Genest, C.; Rémillard, B.; Beaudoin, D. Goodness-of-fit tests for copulas: A review and a power study. *Insur. Math. Econ.* **2009**, *44*, 199–213, doi:10.1016/j.insmatheco.2007.10.005.
64. Salvadori, G.; Durante, F.; De Michele, C. On the return period and design in a multivariate framework. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 3293–3305, doi:10.5194/hess-15-3293-2011.

65. Pappada, R.; Durante, F.; Salvadori, G. Quantification of the environmental structural risk with spoiling ties: Is randomization worthwhile? *Stoch. Environ. Res. Risk Assess.* **2017**, *31*, 2483–2497, doi:10.1007/s00477-016-1357-9.
66. De Michele, C.; Salvadori, G.; Vezzoli, R.; Pecora, S. Multivariate assessment of droughts: Frequency analysis and Dynamic Return Period. *Water Resour. Res.* **2013**, *49*, 6985–6994, doi:10.1002/wrcr.20551.
67. Salvadori, G.; Tomasicchio, G.R.; D'Alessandro, F. Practical guidelines for multivariate analysis and design in coastal and off-shore engineering. *Coast. Eng.* **2014**, *88*, 1–14, doi:10.1016/j.coastaleng.2014.01.011.
68. Brunner, M.I.; Seibert, J.; Favre, A.C. Bivariate return periods and their importance for flood peak and volume estimation. *Wiley Interdiscip. Rev. Water* **2016**, *3*, 819–833, doi:10.1002/wat2.1173.
69. Kroll, C.N.; Song, P. Impact of multicollinearity on small sample hydrologic regression models. *Water Resour. Res.* **2013**, *49*, 3756–3769, doi:10.1002/wrcr.20315.