

Article

Vulnerability Assessment of Dam Water Supply Capacity Based on Bivariate Frequency Analysis Using Copula

Chulsang Yoo  and Eunsam 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: 23 July 2018; Accepted: 16 August 2018; Published: 21 August 2018



Abstract: The vulnerability of the water supply capacity of a dam is defined as the expected deficit volume from a typical water deficit event. In this study, a water deficit event was assumed to be a rectangle composed of the deficit duration and deficit intensity whose occurrence probability was then estimated by the bivariate frequency analysis based on the copula method. This approach is different from the conventional one based on the assumption of the same occurrence probability for all events. This proposed method was applied to the Namgang dam in Korea as an example and the resulting estimate of the vulnerability was compared with the conventional method. First, the ‘OR’ concept was found to be better than the ‘AND’ concept in the calculation of the occurrence probability. Additionally, based on the consideration of multicollinearity, it could be concluded that the occurrence probability should be estimated by considering the water deficit intensity and duration. For the Namgang dam, the vulnerability was determined to be $9.11 \times 10^6 \text{ m}^3$, which is about 3% of the total storage capacity. This estimated vulnerability is also about 70% of the amount estimated by applying the conventional method with the same occurrence probability for all water deficit events.

Keywords: vulnerability assessment; water supply capacity; water deficit event; bivariate frequency analysis; copula model

1. Introduction

Since most of the annual total rainfall in Korea is concentrated in the rainy season from June to September, many multi-purpose dams have been constructed to save water resources. However, due to the increasing demand for domestic, industrial, agricultural, and environmental water use, several river basins are still exposed to the danger of a water shortage [1–5]. Many hydrologists have claimed that additional water resources should be secured by constructing more dams [6–9]. However, it is also argued that any project for securing additional water resources cannot be started before the quantitative evaluation of existing dams [10].

In Korea, the concept of firm yield had been used as one of the main criteria for designing a dam. Based on this concept of firm yield, the water supply capacity was determined by the water budget analysis with the aim of a specific period of maximum drought (generally one or two years) [11]. In the 1970s, the inflow condition recorded during the period of the drought from 1967 to 1968 was the most severe one considered. In the 1980s, the concept of firm yield was still applied to the design of a dam, but the inflow data recorded during the last 20 years could be considered. In the 1990s, the concept of reliability was introduced into Korea.

The reliability may be the most widely used evaluation criterion for the water supply capacity of a dam [12–16]. The term reliability is generally used in Korea and sometimes the volume reliability

and resiliency are estimated as references. These concepts are all easy to understand and easy to apply. However, the concept of vulnerability has never been used because it is rather complicated in its application. The vulnerability may be the most comprehensive measure for the water supply capacity of a dam. In addition, it can provide more detailed information about the water supply capacity such as the water deficit volume.

Generally, the vulnerability indicates the extent to which a system experienced the exposure to a risk factor [17–20]. In a water resource area, the vulnerability is defined as the expected deficit volume from a typical water deficit event [12]. The vulnerability can be calculated as the sum of the volume of all the water deficit events multiplied by their occurrence probabilities. Unfortunately, the occurrence probability of a water deficit event is not easy to estimate. Thus, it is generally assumed to be constant. However, this assumption of constant occurrence probability results in the overestimation of the vulnerability since it counts the extreme events too highly. To estimate the occurrence probability, a multivariate frequency analysis of water deficit events is required. As can be imagined, a water deficit event is quantified by many components like the total deficit volume, deficit duration, peak deficit volume, etc.

This study focuses on the estimation of the occurrence probability of a water deficit event by applying the copula method. For this frequency analysis, a water deficit event should be modeled rather simply but systematically. In this study a water deficit event is assumed to be a rectangle composed of the deficit duration and deficit intensity. The deficit volume is simply the multiplication of deficit duration and deficit intensity. Even though the selection problem of two variables among three remains, the bivariate frequency analysis can be done using the copula. Once the occurrence probability of a water deficit event is estimated, the vulnerability can also be estimated by applying its definition. This study considers the Namgang dam in Korea as an example and the resulting estimate of the vulnerability will be compared with that based on the conventional assumption of constant occurrence probability.

2. Theoretical Background

2.1. Risk Assessment Measures for a Water Supply System

The reliability, resiliency, and vulnerability are regarded as representative water supply safety measures for a water resource system. To explain these three measures, it is important to define a water deficit event. As can be seen in Figure 1, these water supply safety measures are defined by the characteristics of water deficit events. The water deficit event in this study means the condition in which the actual water supply is less than the design of the water supply. The shaded area of Figure 1 indicates an example of a water deficit event. In this figure, $v(j)$, the shaded area indicates the deficit volume for the j -th water deficit event and $d(j)$ indicates the deficit duration for the j -th water deficit event.

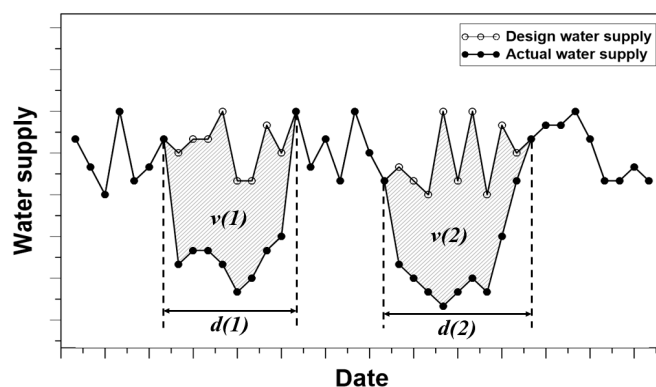


Figure 1. Definition of water deficit volume ($v(i)$) and duration ($d(i)$) (the water deficit intensity is calculated by dividing the deficit volume by the deficit duration).

The reliability may be the oldest measure that has been used to assess the performance of a water resource systems [16]. In general, the reliability refers to the probability that the system will satisfy the design water supply, which can be expressed using the equation below.

$$\alpha = \text{Prob} [X \in S] \quad (1)$$

where X is the state variable of the water resource system expressed as satisfaction or failure and S is the satisfaction state. A way of calculating the reliability is a method of using the ratio of the period in which the water resource system satisfies the design water supply. The reliability calculated in this way is called the temporal reliability. The method of calculating the temporal reliability can be expressed with the equation below.

$$\alpha = 1 - \frac{\sum_{j=1}^M d(j)}{T} \quad (2)$$

where α is the (temporal) reliability, which has the probability value from 0 to 1, M is the number of water deficit events, and T is the entire period to be analyzed. The temporal reliability is simple and convenient to calculate because it only considers the satisfaction or failure of the design water supply for a certain period. However, this measure fails to evaluate the details of the water deficit event. For example, the temporal reliability cannot provide any quantitative information about the degree of damage.

The resiliency is another measure to evaluate how quickly the water resource system returns to the satisfactory state after it has entered the failure state. The resiliency can be expressed as the conditional probability of the satisfactory state at the current time step given the failure state at the previous time step. The resiliency is defined by Equation (3).

$$\begin{aligned} \beta &= \text{Prob} [X_t \in S \mid X_{t-1} \in F] \\ &= \frac{\text{Prob} [X_t \in S, X_{t-1} \in F]}{\text{Prob} [X_{t-1} \in F]} \end{aligned} \quad (3)$$

where β is the resiliency as a conditional probability and X_t is a state variable at time t . To calculate the resiliency, the probability must be calculated that the system is in the failure state (F) at time $t - 1$ and in the satisfactory state (S) at time t . This probability can be calculated by using Z_t , which is defined using the equation below.

$$Z_t \begin{cases} 1, & X_{t-1} \in F, X_t \in S \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

Thus, the probability that the system is in the failure state at time $t - 1$ and in the satisfactory state at time t can be calculated by $\frac{1}{T} \sum_{t=1}^T Z_t$. As the probability of being failed at time $t - 1$ (the denominator of Equation (3)) equals $1 - \alpha$, the degree of resiliency is expressed by using the equation below.

$$\beta = \frac{\sum_{t=1}^T Z_t / T}{1 - \alpha} \quad (5)$$

Lastly, the vulnerability is used to estimate the extent of the deficit volume from a typical water deficit event. Based on the definition of Hashimoto et al. [12], the vulnerability can be calculated as the sum of the volume of all the water deficit events multiplied by their occurrence probabilities. The vulnerability is expressed using the equation below.

$$\gamma = \sum_{j=1}^M v(j) \times e(j) \quad (6)$$

where γ is the vulnerability, which has the same unit of water deficit volume (m^3), and $e(j)$ is the probability of occurrence of the j -th water deficit event. In this scenario, it should be remembered that the sum of $e(j)$ is equal to one, i.e., $\sum_{j=1}^M e(j) = 1$.

In real practice, the reliability has been most frequently used since it is easy to calculate and easy to understand. On the other hand, neither the resiliency nor the vulnerability have been used that much. The resiliency shows just one aspect of the water resource system similar to the reliability. Even though the vulnerability considers both the volume and occurrence probability of water deficit events, the calculation of the occurrence probability has been a difficult problem. As a simplification, the occurrence probability has been generally assumed as $e(j) = 1/M$ [15,21,22]. Thus, the estimated vulnerability cannot be the true value of representing the given water resource system.

This study focuses on the vulnerability of a water resource system. Since a water deficit event is defined by its several components, like the deficit volume, deficit duration, etc., the calculation of its occurrence probability cannot be based on univariate frequency analysis. In this study, the copula method was introduced to solve this problem of estimating the occurrence probability. The vulnerability analysis result from a copula-based method will become more relevant information for the management of existing dam and decision-making of new dam construction.

2.2. Bivariate Frequency Analysis Using the Copula

2.2.1. Basic Concept of the Copula

Sklar [23] introduced the copula as a mathematical concept for the formation of a multivariate distribution function. The copula has been applied to derive one joint probability density function based on various PDFs [24–26]. By applying copula, any kind of PDF can be combined to one joint PDF. As a result, joint PDF representing the characteristics of various PDFs can be derived. The fundamental equation explaining the copula can be expressed by using the equation below.

$$H(x_1, x_2, \dots, x_n) = C[F_1(x_1), F_2(x_2), \dots, F_n(x_n)] \quad (7)$$

where H is the cumulative probability that is calculated using the joint PDF, C is the copula, and $F_i(x_i)$ is the cumulative probability that is calculated using the i -th PDF. Equation (7) shows that the copula can be used to combine n PDFs into a single joint PDF.

There are three major copula families: the Archimedean copula, the elliptical copula, and the Marshall–Olkin copula [27]. In this study, the Archimedean and elliptical copula families were selected for analysis. The reason for the selection is that copula belongs to these two families, which is easy to construct and is known for its wide applicability [24]. A total of four copula models were considered in this study including the Clayton model [28], the Frank model [29], the Gumbel–Hougaard model [30], and the Gaussian model [31]. Among them, the Clayton model, the Frank model, and the Gumbel–Hougaard model are classified as the Archimedean copula. Additionally, the Gaussian model is classified as the elliptical copula.

Table 1 summarizes the joint PDFs of these copula models. In Table 1, θ is the parameter for copula model. u_1 and u_2 are the cumulative probabilities calculated by the marginal PDF of each two target variables [32]. Φ is the cumulative distribution function of the standard gaussian distribution.

Table 1. Joint PDF of copula models considered in this study.

Copula Model	Joint PDF
Clayton	$C(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}$
Frank	$C(u_1, u_2) = -\frac{1}{\theta} \ln[1 + \frac{\exp(-\theta u_1) - 1}{\exp(-\theta) - 1} \frac{\exp(-\theta u_2) - 1}{\exp(-\theta) - 1}]$
Gumbel–Hougaard	$C(u_1, u_2) = \exp(-((-\ln(u_1))^\theta + (-\ln(u_2))^\theta)^{1/\theta})$
Gaussian	$C(u_1, u_2) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\theta^2}} \exp(-\frac{s^2+w^2-2sw\theta}{2(1-\theta^2)}) dsdw$

2.2.2. Parameter Estimation

In this study, Kendall’s τ , which is known as the rank correlation of the data, was used to calculate the parameter of the copula model. The parameter of the copula model in this study can be easily estimated using the Kendall’s τ because this study considered the copula models with only one parameter. The equation for Kendall’s τ of the two target variables can be expressed with the equation below [25].

$$\tau = \frac{P_n - Q_n}{nC_2} \tag{8}$$

where n is the number of data, (X_i, Y_i) and (X_j, Y_j) are data pairs for $i \neq j$, P_n is the number of cases where $(X_i - X_j)(Y_i - Y_j) > 0$, Q_n is the number of cases where $(X_i - X_j)(Y_i - Y_j) < 0$, and nC_2 is the number of combinations for the selection of the two variables among n elements.

The parameter θ of the copula, then, was estimated by using the relationship between parameter θ and Kendall’s τ . Table 2 summarizes the relationships for the copula models that are considered in this study.

Table 2. Relationship between the Kendall’s τ and the parameter θ .

Copula Model	Relationship between Kendall’s τ and the Parameter θ
Clayton	$\tau = \frac{\theta}{\theta+2}$
Frank	$\tau = 1 - \frac{4}{\theta} + 4 \frac{D_1(\theta)}{\theta}, D_1(\theta) = \int_0^\theta \frac{t}{e^t-1} dt$
Gumbel–Hougaard	$\tau = 1 - \frac{1}{\theta}$
Gaussian	$\tau = \frac{2}{\pi} \arcsin\theta$

Each copula model has a valid range for the parameter θ and only the parameter θ estimated in that range is significant. The valid range for the Clayton model is $[0, \infty)$ and the valid range for the Gumbel–Hougaard model is $[1, \infty)$. In the case of the Gaussian model, the valid range is $[-1, 1]$. For the Frank model, the valid range is $(-\infty, \infty)$ except for the origin [24].

2.2.3. Determination of the Optimal Copula Model

Among four copula models considered in this study, the optimal copula model was determined based on an Akaike information criterion (AIC) test statistic [33]. Although the AIC is not appropriate for the goodness-of-fit test, it could be used to determine the best model among many options. For example, Zhang and Singh [32] explained how to select the best copula model among many options by comparing AIC of each copula.

The AIC test statistic quantifies the difference between the empirical cumulative probability and the cumulative probability calculated from the copula model. The empirical cumulative probability

for given data can be calculated using the Gringorten plotting position formula [34]. The Gringorten plotting position formula can be expressed by using the equation below.

$$H(x_i, y_i) = P(X \leq x_i, Y \leq y_i) = \frac{\text{Number of } (x \leq x_i \text{ and } y \leq y_i) - 0.44}{n + 0.12} \quad (9)$$

In Equation (9), $H(x_i, y_i)$ is the empirical cumulative probability of the data (x_i, y_i) .

The difference between the empirical cumulative probability and the cumulative probability of the copula model is quantified with the mean squared error (MSE) [35]. The MSE for AIC is defined using the formula below.

$$MSE = \frac{1}{n} \sum_{i=1}^n \{x_e(i) - x_o(i)\}^2 \quad (10)$$

In Equation (10), x_e is the cumulative probability from the copula model and x_o is the empirical cumulative probability. Lastly, the AIC test statistic can be calculated using the formula below.

$$AIC = n \log(MSE) + 2(\text{number of fitted parameters}) \quad (11)$$

In this case, the copula model with the lower AIC test statistic is evaluated as the better copula model for the given data. Therefore, the optimal copula model for the given data is the model with the lowest AIC value [33].

However, having the lowest AIC does not guarantee the absolute goodness-of-fit of the copula model. To verify the absolute goodness-of-fit of the copula model, the Kolmogorov–Smirnov (K–S) test was conducted on the optimal copula. In the K–S test, the goodness-of-fit of the model is evaluated based on the maximum absolute difference between the empirical cumulative probability and the cumulative probability calculated from the copula [36]. The absolute goodness-of-fit of the optimal copula model is examined by comparing the difference with the critical value of the K–S test at a certain significance level. In this study, the significance level of 5% was used for the critical value of the K–S test.

2.2.4. Occurrence Probability Calculated by the Copula Model

First, the occurrence probability can be calculated by considering the case where one or two variables exceed the threshold value. This occurrence probability is expressed as P_{or} . In addition, the occurrence probability can be defined by the probability that both variables exceed the threshold value. This case is generally notated by P_{and} . The P_{or} and P_{and} for the data (x_i, y_i) can be calculated using the following equations.

$$P_{or} = 1 - F_{X,Y}(x_i, y_i) \quad (12)$$

$$P_{and} = 1 - F_X(x_i) - F_Y(y_i) + F_{X,Y}(x_i, y_i) \quad (13)$$

where $F_{X,Y}(x_i, y_i)$ is the joint cumulative probability of the copula model and $F_X(x_i)$ and $F_Y(y_i)$ are the cumulative probabilities of x_i and y_i , respectively.

3. Study Area and Data

The Nam river is a major tributary of the Nakdong river, which covers the southeastern part of the Korean Peninsula. The Namgang dam is located 80 km from the confluence point of the Nakdong river (Figure 2a). Figure 2b shows the river network and outlet of the Namgang dam basin and Figure 2c is the Digital Elevation Model (DEM) of the Namgang dam basin (Figure 2c) where the highest point is marked as the “Jiri mountain.” The basin area of the Namgang dam is 3466 km² and the main channel length is 10⁸ km. Even though it has a large basin area, the total storage capacity (3.1×10^8 m³) and annual water supply capacity (5.7×10^8 m³) of the Namgang dam are relatively small. For reference, the basin area of the Kurobe dam is 188.5 km², which is about 5% of that for the Namgang dam.

However, the total storage capacity of the Kurobe dam is about $1.9 \times 10^8 \text{ m}^3$, which is about 61% of that for the Namgang dam.

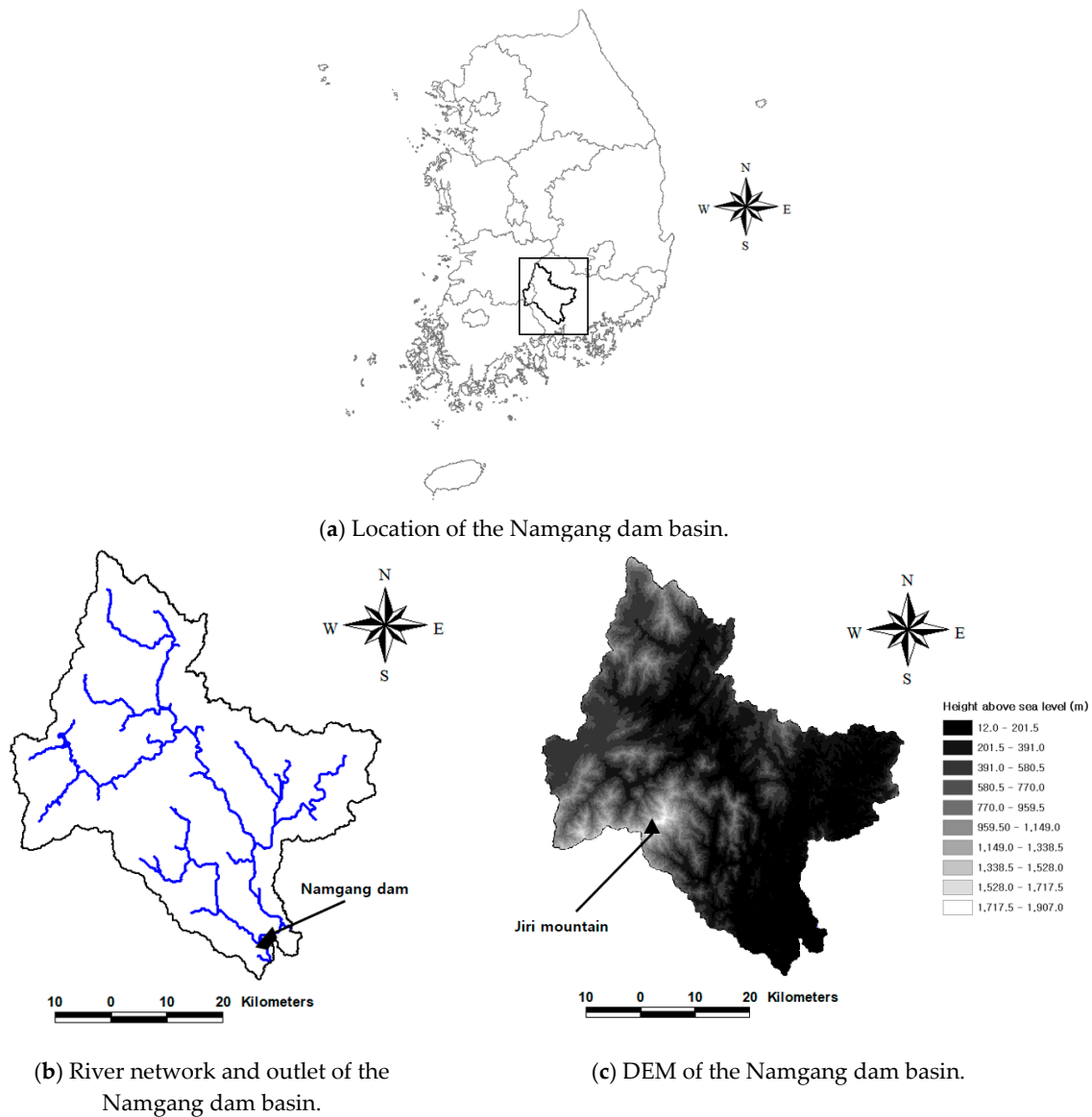


Figure 2. Location of the Namgang dam basin and its characteristics.

This study used the daily water supply time series from the Namgang dam from 1981 to 2012. The water deficit event is defined as the consecutive day in a water deficit condition when the dam water supply is less than the design water supply. In this study, all the water deficit events were considered for further analysis and a total of 45 water deficit events were collected.

A water deficit event can be quantified by the deficit duration, deficit volume, mean deficit volume (or, deficit severity), peak deficit volume, and more. Since the structure of a water deficit event is rather complicated, a simplified structure is generally introduced. The simplest way of representing a water deficit event may be to use a rectangular pulse. The rectangular pulse has been widely used for modeling rainfall events [37–42], as well as drought events [43]. This study also adopted this simplification to represent a water deficit event with its deficit duration, deficit volume, and deficit intensity (i.e., the deficit volume divided by its deficit duration).

Table 3 summarizes the basic statistics (mean, standard deviation, and range) of these components of the water deficit events considered in this study. The means of the deficit duration, deficit volume, and deficit intensity were calculated as 15.0 day, $13.5 \times 10^6 \text{ m}^3$, and $12.3 \times 10^6 \text{ m}^3/\text{day}$, respectively. The standard deviations are all proportional to the means with their coefficient of variations of approximately (0.57, 0.66, and 1.25) for the deficit duration, deficit volume, and deficit intensity, respectively. The ranges of those components were also summarized as (1 to 161) day for the deficit duration, $(0.1 \text{ to } 118.9) \times 10^6 \text{ m}^3$ for the deficit volume, and $(0.1 \text{ to } 2.7) \times 10^6 \text{ m}^3/\text{day}$ for the deficit intensity.

Table 3. The mean, standard deviation, and ranges of three components of the water deficit events observed at the Namgang dam.

Variable	Mean	Standard Deviation	Range
Deficit duration (day)	15.0	26.5	1–161
Deficit volume (10^6 m^3)	13.5	20.4	0.1–118.9
Deficit intensity ($10^6 \text{ m}^3/\text{day}$)	1.0	0.8	0.1–2.7

Figure 3 shows the box-plot for the basic components of water deficit events observed at the Namgang dam. In Figure 3, each of the outliers outside the 1.5 inter quartile range is marked by the 'X' symbol. There were three outliers for the deficit duration (Figure 3a) including one for the deficit volume (Figure 3b) and no outlier was found for the deficit intensity (Figure 3c). All these outliers are from three water deficit events among which one event was extreme. The effect of these three events was also found to be big enough to severely distort the basic statistics. Therefore, in this study, they were excluded in the fitting procedure of the PDFs. After selecting the PDFs, the goodness-of-fit test was also repeated with all the water deficit events including those three rare events. The vulnerability assessment was also conducted with all the water deficit events including those three rare events.

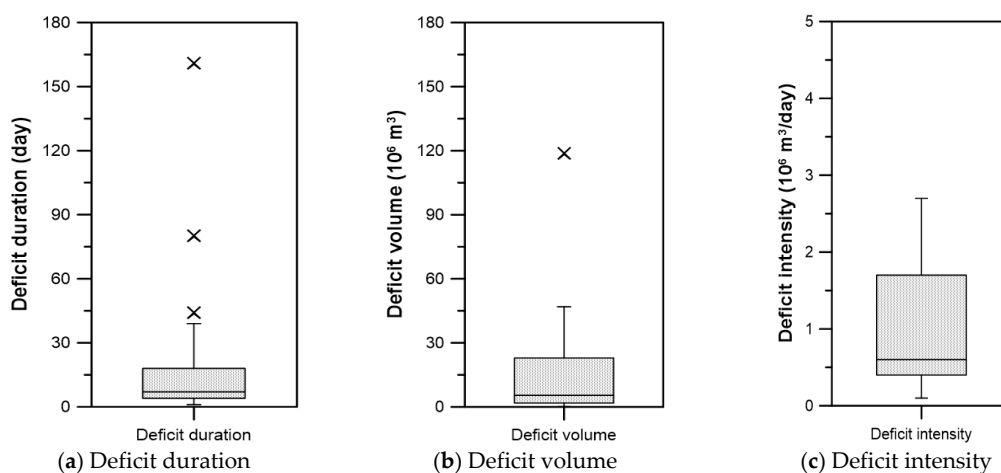


Figure 3. Box-plots of three components of water deficit events observed at the Namgang dam.

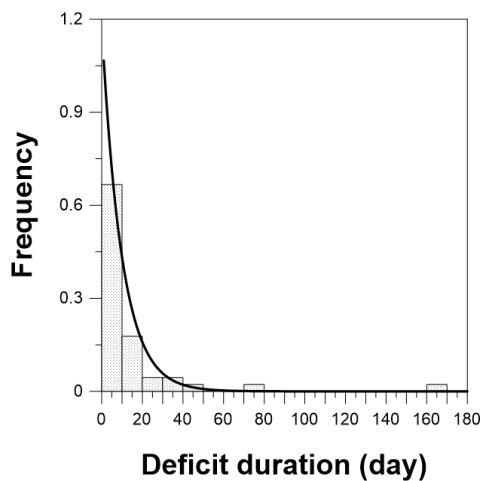
4. Results

4.1. Probability Density Functions

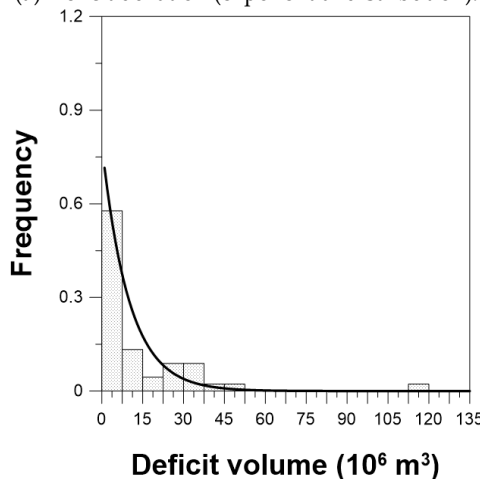
To calculate the occurrence probability using the copula model, the PDF for each variable of the water deficit event must be determined. In this study, the exponential distribution was chosen for the deficit duration and deficit volume and the log-Gumbel distribution was chosen for the deficit intensity. The reason for selecting the exponential distribution is that the occurrence characteristic of

water deficit events is known to follow the Poisson process. In addition, it was also considered that the deficit duration and deficit volume data were positively skewed and more than 50% of the data were distributed near zero. On the other hand, the log-Gumbel distribution for the deficit intensity was determined based on the comparison of the data histogram and several PDFs. In this study, normal distribution, log-normal distribution, gamma distribution, generalized extreme value distribution, Gumbel distribution, and log-Gumbel distribution were considered as options.

For parameter estimation of the PDF, the maximum likelihood method was applied. Figure 4 compares the histogram and the PDF for each component of the water deficit event. Each PDF in Figure 4 was tested for goodness-of-fit by the χ^2 test. The exponential distribution for the deficit duration was found to have 0.19 of the χ^2 statistic, which is far smaller than the criterion for the 5% significance level. The exponential distribution for the deficit volume was also found to have a smaller χ^2 statistic of 1.34. In the case of the log-Gumbel distribution for deficit intensity, the χ^2 statistic was calculated as 4.29. This is larger than that for the deficit duration and deficit volume but still smaller than the criterion for the 5% significance level. In conclusion, all distributions were found to be valid under the 5% significance level. As mentioned earlier, all the water deficit events were considered in this goodness-of-fit test.



(a) Deficit duration (exponential distribution).



(b) Deficit volume (exponential distribution).

Figure 4. Cont.

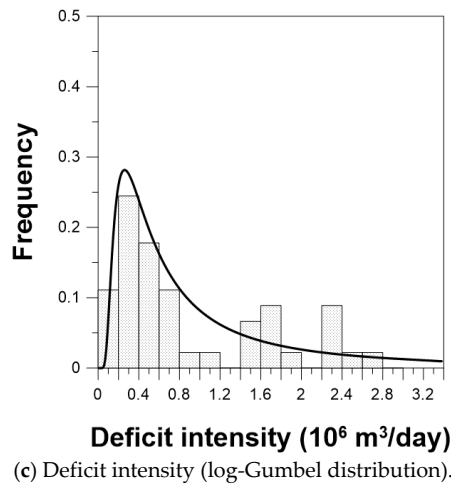


Figure 4. Comparison of data histograms and PDFs for three components of water deficit events observed at the Namgang dam.

4.2. Optimal Copula Model for the Water Deficit Events

The Kendall’s τ is necessary for estimating the parameter of the copula models that are considered in this study. The Kendall’s τ was estimated as 0.67 for the deficit volume and the deficit duration, 0.26 was estimated for the deficit intensity and the rainfall duration, and 0.52 was estimated for the deficit volume and the deficit intensity. For all bivariate data pairs, the positive correlation was estimated. Among the three, the strength of correlation for the deficit volume and the deficit duration was highest while that for the deficit volume and the deficit intensity was lowest. Figure 5 shows the scatter plot of three bivariate data pairs excluding the three outliers.

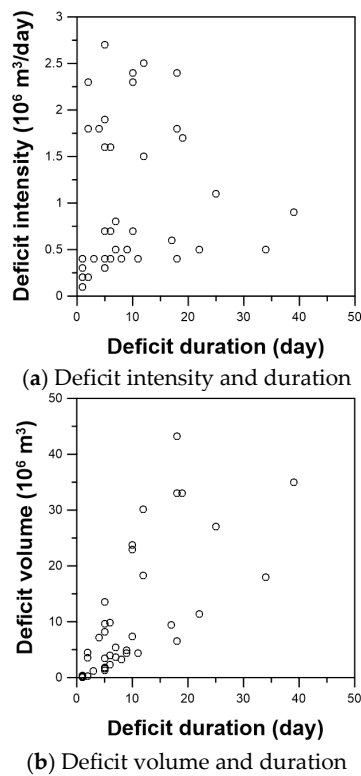


Figure 5. Cont.

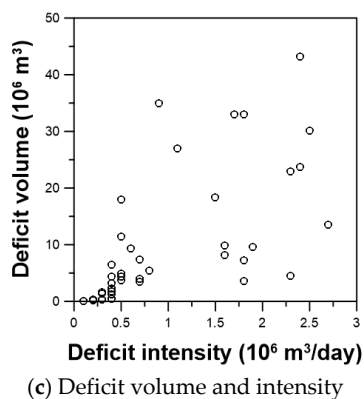


Figure 5. Scatter plots for the three pairs of components of water deficit events observed at the Namgang dam (a) deficit intensity and duration; (b) deficit volume and duration; (c) deficit volume and intensity.

The parameter θ of the copula model was then estimated using the relationship between the Kendall's τ and the parameter θ , which is given in Table 2. The parameter θ for the Clayton, Frank, Gumbel-Hougaard, and Gaussian copula models, respectively, were determined, which is shown in Table 4. Every parameter θ was found to be in the valid range of the four copula models.

Table 4. The parameter θ estimated for each data pair with different copulas.

Copula Model	Deficit Volume and Duration	Deficit Intensity and Duration	Deficit Volume and Intensity
Clayton	3.97	0.70	2.13
Frank	11.22	5.36	9.01
Gumbel-Hougaard	2.99	1.35	2.06
Gaussian	0.87	0.40	0.72

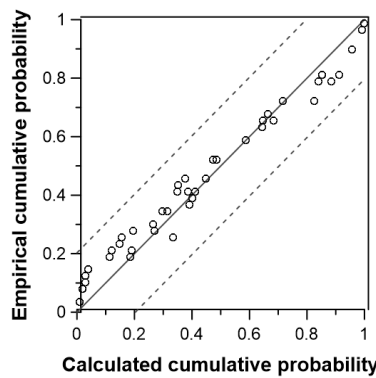
The *MSE* and the *AIC* were estimated for selecting the optimal copula model for each case of the bivariate data pair (Table 5). Since the *MSE* and *AIC* values are all penalty measures, the lower value indicates the more favorable models. As a result, the Clayton model was selected as the optimal copula model for the deficit volume and deficit duration, the Gaussian model was selected for the deficit intensity and the deficit duration, and the Frank model was selected for the deficit volume and the deficit intensity.

Table 5. The *MSEs* and *AICs* calculated for three bivariate data pairs of the water deficit events observed at the Namgang dam.

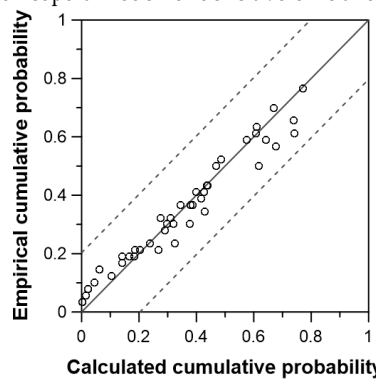
Data Pair		Clayton	Frank	Gumbel-Hougaard	Gaussian
<i>MSE</i>	Deficit volume and duration	0.00351	0.00398	0.00404	0.00440
	Deficit intensity and duration	0.0287	0.00397	0.0313	0.0313
	Deficit volume and intensity	0.0349	0.00401	0.0378	0.0378
<i>AIC</i>	Deficit volume and duration	−108.488	−106.005	−104.051	−105.728
	Deficit intensity and duration	−67.370	−106.052	−65.703	−114.095
	Deficit volume and intensity	−63.599	−105.839	−62.000	−102.689

Figure 6 compares the empirical cumulative probability and cumulative probability calculated by the optimal copula model. In this figure, the *x*-axis represents the cumulative probability calculated by the optimal copula while the *y*-axis represents the empirical cumulative probability. The dotted lines in this figure indicate the critical value of the K-S test based on the 5% significance level. As can be seen from Figure 6, the discrepancies for all optimal copulas lie within the upper and lower limits of

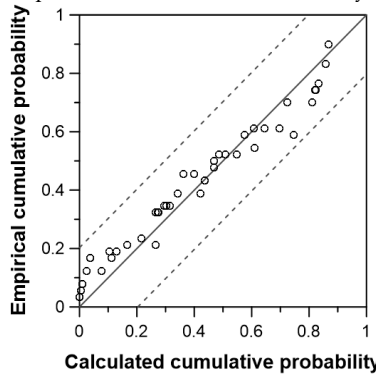
the 5% significance level (0.203 for $n = 45$ and $\alpha = 0.05$). The goodness-of-fit for all optimal copulas were verified under the 5% significance level.



(a) Clayton copula model for deficit volume and duration.



(b) Gaussian copula model for the deficit intensity and duration.



(c) Frank copula model for the deficit volume and intensity.

Figure 6. K-plots and K-S test criteria for three optimal copula models considered for the water deficit events observed at the Namgang dam.

4.3. Occurrence Probability

Using the copula model determined for each pair of two variables constituting the water deficit event, the occurrence probability of each water deficit event was estimated. Both the 'AND' and 'OR' concepts were applied in this calculation. Figure 7 compares the estimated occurrence probabilities of each water deficit event for both the 'AND' and 'OR' concepts. The occurrence probability in Figure 7 is divided by the sum of the exceedance probabilities of all the water deficit events estimated by applying the determined copula model to satisfy the concept of vulnerability. The mean value of the exceedance probabilities of all the deficit events was 0.621 in the 'AND' case and 0.357 in the 'OR' case. The dotted line in Figure 7 indicates the occurrence probability considered in the conventional method. Since the number of events is 45, the occurrence probability is calculated to be $1/45 = 0.022$.

Figure 7 clearly shows the difference between the ‘AND’ and ‘OR’ cases. Since the occurrence probability in the ‘AND’ case is calculated by way of multiplication of both the occurrence probabilities of two variables, the resulting occurrence probability of a water deficit event can be much smaller or much bigger depending on the occurrence probability of each variable. Exactly the opposite behavior can be expected for the ‘OR’ case. In this case, the occurrence probability of a water deficit event is calculated as a way of summation and it is generally higher than that of one variable. For example, when considering the deficit volume and duration, in the ‘AND’ case, the occurrence probability of a water deficit event with deficit volume $35.0 \times 10^6 \text{ m}^3$ that occurred in 1977 was estimated to be 0.00011. On the other hand, in the ‘OR’ case, the occurrence probability of the same event was calculated to be 0.0016, which is obviously much higher than that in the ‘AND’ case.

As can be seen from Figure 7a, the variation of estimated occurrence probabilities in the ‘AND’ case became wide enough to reach 0.068. On the other hand, as can be seen from Figure 7b, the variation of the estimated occurrence probabilities in the ‘OR’ case was much smaller than that in the ‘AND’ case. The maximum occurrence probability was estimated to be only 0.037. As a result, in the calculation of the vulnerability, the role of extreme events becomes minimal in the ‘AND’ case but can be much bigger in the ‘OR’ case. However, at this point, it is not clear if either of the ‘AND’ or ‘OR’ case is appropriate for the estimation of the vulnerability, which will be discussed in the following section.

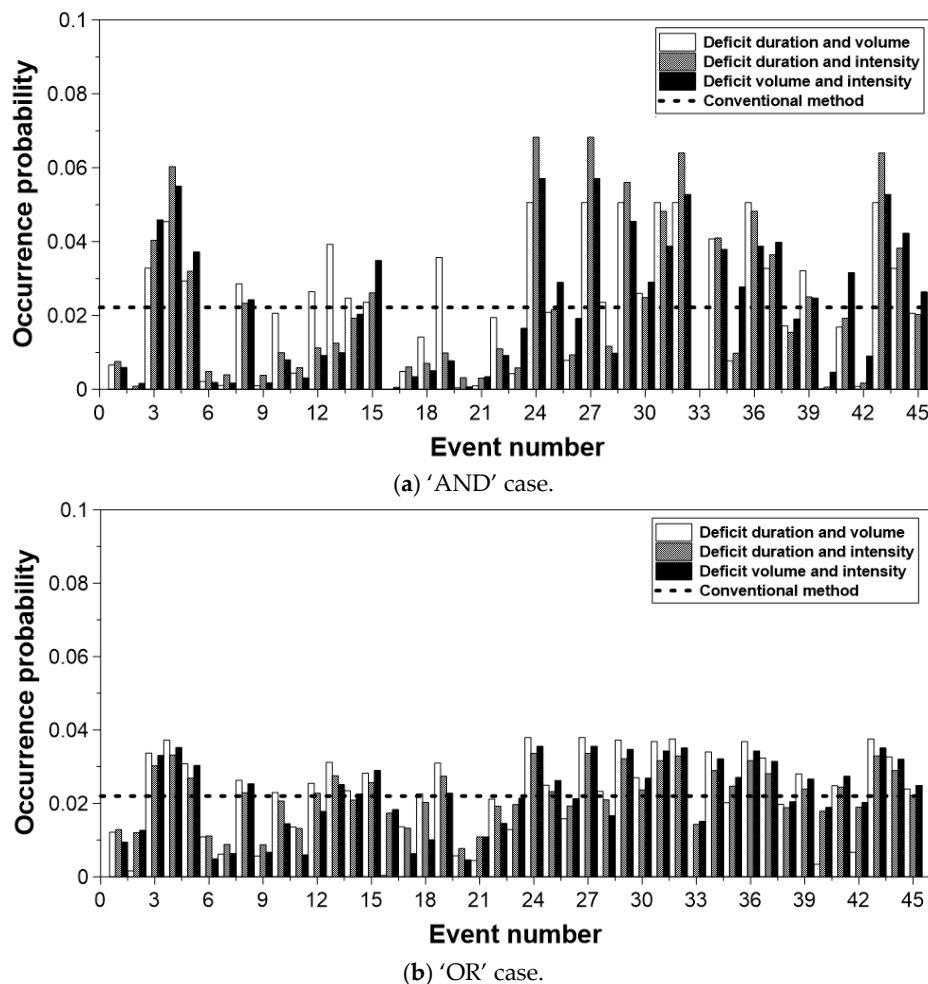


Figure 7. Comparison of normalized occurrence probabilities estimated for three different pairs of two variables constituting the water deficit events observed at the Namgang dam including for (a) the ‘AND’ case and (b) the ‘OR’ case.

It was also found that the selection of the data pair constituting the water deficit events has a significant influence on the results of the occurrence probability. Theoretically, the occurrence probabilities calculated from the same deficit event have to be identical. However, due to the uncertainty in the process of estimating the parameter of the copula and calculating the occurrence probability, the three occurrence probability results show important differences. For example, in the 'AND' case, the occurrence probability of the water deficit event with a deficit volume of $33.0 \times 10^6 \text{ m}^3$, which occurred in 1978, was 0.00098 when considering the deficit volume and duration but 0.038 when considering the deficit intensity and duration and 0.0017 when considering the deficit volume and intensity. This problem was a little alleviated in the 'OR' case even though the difference was still significant. For example, the occurrence probability of the water deficit event of the same event in 1978 was 0.0056 when considering the deficit volume and duration, but it was 0.0087 when considering the deficit intensity and duration and 0.0067 when considering the deficit volume and intensity.

4.4. Vulnerability

Table 6 compares the vulnerabilities estimated for the three different cases of selecting the data pair and two different cases of estimating the occurrence probability. As mentioned in the previous section, the difference between 'AND' and 'OR' cases when estimating the occurrence probability was also very clear in the calculation of the vulnerability. In the case of considering the 'AND' concept, the occurrence probability of an extreme event becomes extreme. As a result, the effect of these extreme events on the vulnerability estimation becomes very limited. Due to this reason, the vulnerability was estimated to be very small in the 'AND' case and the difference among the three cases of selecting the data pair was also small. The smallest one was just $3.16 \times 10^6 \text{ m}^3$ for the data pair of the deficit intensity and duration while the largest one was $3.38 \times 10^6 \text{ m}^3$ for the data pair of the deficit volume and intensity. These estimated vulnerabilities are all on the level of 25% of that based on the conventional method of using the same occurrence probability (i.e., $13.54 \times 10^6 \text{ m}^3$).

On the other hand, the 'OR' case resulted in quite reasonable results of vulnerability estimation. Since the extreme events were all significantly considered in estimating the vulnerability, the vulnerability was estimated to be much higher than that in the 'AND' case. Additionally, depending on the selected data pair, the estimated vulnerability was significantly different. It was $5.34 \times 10^6 \text{ m}^3$ for the data pair of the deficit volume and duration, but it was $9.11 \times 10^6 \text{ m}^3$ for the data pair of the deficit intensity and duration. These estimated vulnerabilities are at the level of (40% to 70%) of that based on the conventional method.

Table 6. Vulnerabilities estimated for three different pairs of two variables and two cases of calculating the occurrence probability, i.e., 'AND' and 'OR' cases.

Data Pair	'AND' Case	'OR' Case
Deficit volume and duration	$3.27 \times 10^6 \text{ m}^3$	$5.34 \times 10^6 \text{ m}^3$
Deficit intensity and duration	$3.16 \times 10^6 \text{ m}^3$	$9.11 \times 10^6 \text{ m}^3$
Deficit volume and intensity	$3.38 \times 10^6 \text{ m}^3$	$8.30 \times 10^6 \text{ m}^3$
Conventional method	$13.54 \times 10^6 \text{ m}^3$	

Lastly, it is necessary to select one of the three results summarized in Table 6. The vulnerability is estimated by considering that the deficit volume and duration is $5.34 \times 10^6 \text{ m}^3$, that the deficit intensity and duration is $9.11 \times 10^6 \text{ m}^3$, and that the deficit volume and intensity is $8.30 \times 10^6 \text{ m}^3$. Since the vulnerability was calculated differently for the three different pairs of the two variables considered even though any pair of the two variables defined the same water deficit event, it is necessary to select any one of them for further remediation of the dam water supply plan. For this selection problem, the concept of multicollinearity can be used as an indicator.

The multicollinearity is defined as the lack of independence or the presence of interdependence among independent variables in the regression analysis [44,45]. The interdependent and independent

variables can lead to low quality of the resulting parameter estimates [46,47]. Additionally, the effect of the multicollinearity distortion increases with the increasing degree of correlation [48,49]. This problem is generally evaluated by quantifying the variance inflation factor from a variable [50,51]. The variance inflation factor is used as a measure to quantify the level of the multicollinearity in a multiple regression analysis.

Furthermore, the above results in Table 6 could be expected by comparing the variance inflation factors estimated for the three components considered in this study. The variance inflation factor of the deficit duration was estimated as 3.46 and the deficit volume was estimated to be 5.87. In the case of the deficit intensity, the variance inflation factor is 2.89. It is generally assumed that, in the case that the variance inflation factor is higher than five, the multicollinearity problem becomes serious [52–55]. Based on this result, it could be concluded that the occurrence probability considering the deficit intensity and duration (i.e., the case excluding the deficit volume) is more reasonable than the other two cases. The vulnerability estimated by considering the deficit intensity and duration, $9.11 \times 10^6 \text{ m}^3$, becomes the most reasonable value for the Namgang dam.

5. Conclusions

This study proposed a method to estimate the occurrence probability of a water deficit event by applying the copula method. In this study, a water deficit event was assumed to be a rectangle composed of the deficit duration and deficit intensity. The deficit volume is simply the product of the deficit duration and deficit intensity. By selecting two variables among three, the bivariate frequency analysis was performed using the copula. The vulnerability was then estimated by multiplying the water deficit volume by its occurrence probability. This study considered the Namgang dam as an example and the resulting estimate of the vulnerability was compared with the conventional assumption of constant occurrence probability. The results are summarized below.

First, the exponential distribution was selected for the deficit duration and deficit volume and the log-Gumbel distribution was selected for the deficit intensity. The goodness-of-fit tested by the test has also validated the use of these PDFs. Additionally, based on the *MSE* and the *AIC*, the Clayton model was selected as the optimal copula model for the deficit volume and deficit duration. The Gaussian model was selected for the deficit intensity and the deficit duration and the Frank model was chosen for the deficit volume and the deficit intensity. The K-S test confirmed these selections of the copula model.

Second, the occurrence probability of each water deficit event was estimated by using the copula model determined for each pair of two variables constituting the water deficit event. Both the 'AND' and 'OR' concepts were applied in this calculation and they were found to be very different from each other. In particular, the occurrence probability in the 'AND' case was estimated to be much smaller or much larger (i.e., wide variation) than that in the 'OR' case. It was also found that the selection of the data pair constituting the water deficit events had a significant influence on the occurrence probability. This problem was serious in the 'AND' case but was rather small in the 'OR' case even though the difference was still significant.

Third, the difference between the 'AND' and 'OR' cases when estimating the occurrence probability was also considered in calculating the vulnerability. In considering the 'AND' concept, the occurrence probability of an extreme event becomes extreme (i.e., a small occurrence probability). As a result, the effect of these extreme events on the vulnerability estimation becomes very limited. Due to this reason, the vulnerability was estimated to be very small in the 'AND' case and the differences among the three cases of selecting the data pair were also small. On the other hand, the 'OR' case resulted in quite reasonable results of vulnerability estimation. Since the extreme events were all significantly considered in estimating the vulnerability, the vulnerability was assessed to be much higher than that in the 'AND' case. The estimated vulnerabilities in the 'AND' case were on the level of 25%, but, in the 'OR' case, they were about 40% to 70% of that based on the conventional method.

Lastly, based on the concept of multicollinearity, it could be concluded that the occurrence probability considering the deficit intensity and duration is more reasonable than the other two cases.

This conclusion was made by considering the variance inflation factor, which is a measure to evaluate the effect of the multicollinearity of a variable. The variance inflation factor was estimated as 2.89 for the deficit intensity, 3.46 for the deficit duration, and 5.87 for the deficit volume. It is generally assumed that the multicollinearity problem becomes serious in the case that the variance inflation factor is higher than five. As a final result in the application of the Namgang dam, the vulnerability was determined to be $9.11 \times 10^6 \text{ m}^3$, which is about 3% of the total storage capacity of $3.1 \times 10^8 \text{ m}^3$. This estimated vulnerability is also about 70% of that estimated by applying the same occurrence probability to all water deficit events.

The vulnerability analysis result from the copula-based method will become more relevant for the management of existing dam and decision-making of new dam construction. In addition, since this method is based on the copula model, any kind of marginal distribution for an analysis variable can be applied. Therefore, a vulnerability assessment method from this study can be broadly applied on a water-deficit event with various types of variables.

More case studies with various types of water resource systems need to be conducted to generalize the vulnerability assessment method in this study. Therefore, for further study, several vulnerabilities results from the global/domestic basin will be analyzed and compared.

Author Contributions: C.Y. conceived and designed the idea of this research and wrote the manuscript. E.C. collected the data and conducted bivariate analysis using that data.

Funding: This work is supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure, and Transport (Grant 18AWMP-B083066-05) and the international cooperation program managed by the National Research Foundation of Korea (2016K2A9A1A06922023).

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

References

- Choe, J.-H.; Heo, E.-N.; Sim, M.-P. The economic impacts of water supply constraints during a drought using input-output analysis. *J. Korea Water Resour. Assoc.* **2000**, *33*, 647–658.
- Chang, H.; Kwon, W.-T. Spatial variations of summer precipitation trends in South Korea, 1973–2005. *Environ. Res. Lett.* **2007**, *2*, 045012. [[CrossRef](#)]
- Seo, S.-S.; Kim, D.-G.; Lee, K.-H.; Kim, H.-S.; Kim, T.-W. Estimation of drought damage based on agricultural and domestic water use. *J. Wetl. Res.* **2009**, *11*, 77–87.
- Lee, G.-M.; Yi, J. Analysis of emergency water supply effects of multipurpose dams using water shortage index. *J. Korea Water Resour. Assoc.* **2012**, *45*, 1143–1156. [[CrossRef](#)]
- Lee, G.-M. Water supply performance assessment of multipurpose dams using sustainability index. *J. Korea Water Resour. Assoc.* **2014**, *47*, 411–420. [[CrossRef](#)]
- Yi, J.-E. Development of optimal operation rule for multipurpose reservoirs system. *J. Korea Water Resour. Assoc.* **2004**, *37*, 487–497. [[CrossRef](#)]
- Yi, J.-E.; Kwon, D.-S. Effects on conservation and flood control systems according in normal water level change from Daechung multi-purpose reservoir. *J. Korea Water Resour. Assoc.* **2007**, *40*, 1–10. [[CrossRef](#)]
- Long-Term Master Plan for Dam Construction (2012–2021)*; Ministry of Land, Transport and Maritime Affairs: Seoul, Korea, 2012.
- Long-Term Comprehensive Water Resource Plans (2001–2020)*; Ministry of Land, Transport and Maritime Affairs: Seoul, Korea, 2016.
- Harada, J.; Yasuda, N. Conservation and improvement of the environment in dam reservoirs. *Int. J. Water Resour. Dev.* **2004**, *20*, 77–96. [[CrossRef](#)]
- Yi, J.; Lee, G.; Cha, K. Idea on the improvement of water yield evaluation methodology. *Mag. Korea Water Resour. Assoc.* **2012**, *45*, 51–57. [[CrossRef](#)]
- Hashimoto, T.; Stedinger, J.R.; Loucks, D.P. Reliability, resiliency, and vulnerability criteria for water resource system performance evaluation. *Water Resour. Res.* **1982**, *18*, 14–20. [[CrossRef](#)]
- Moy, W.S.; Cohon, J.L.; ReVelle, C.S. A programming model for analysis of the reliability, resilience, and vulnerability of a water supply reservoir. *Water Resour. Res.* **1986**, *22*, 489–498. [[CrossRef](#)]

14. Grigg, N.S. Management framework for large-scale water problems. *J. Water Resour. Plan. Manag.* **1996**, *122*, 296–300. [[CrossRef](#)]
15. Kjeldsen, T.R.; Rosbjerg, D. Choice of reliability, resilience and vulnerability estimators for risk assessments of water resources systems/choix d'estimateurs de fiabilité, de résilience et de vulnérabilité pour les analyses de risque de systèmes de ressources en eau. *Hydrol. Sci. J.* **2004**, *49*. [[CrossRef](#)]
16. Lee, G.-M. On the water yield assessment index in water resources system. *Water Future* **2012**, *45*, 71–77.
17. Füssel, H.-M.; Klein, R.J. Climate change vulnerability assessments: An evolution of conceptual thinking. *Clim. Change* **2006**, *75*, 301–329. [[CrossRef](#)]
18. Gain, A.K.; Giupponi, C.; Renaud, F.G. Climate change adaptation and vulnerability assessment of water resources systems in developing countries: A generalized framework and a feasibility study in bangladesh. *Water* **2012**, *4*, 345–366. [[CrossRef](#)]
19. Wu, G.; Li, L.; Ahmad, S.; Chen, X.; Pan, X. A dynamic model for vulnerability assessment of regional water resources in arid areas: a case study of Bayingolin, China. *Water Resour. Manag.* **2013**, *27*, 3085–3101. [[CrossRef](#)]
20. Nicu, I.C. Cultural heritage assessment and vulnerability using analytic hierarchy process and geographic information systems (Valea Oii catchment, North-eastern Romania). An approach to historical maps. *Int. J. Disaster Risk Reduct.* **2016**, *20*, 103–111. [[CrossRef](#)]
21. Jinno, K. Risk assessment of a water supply system during drought. *Int. J. Water Resour. Dev.* **1995**, *11*, 185–204. [[CrossRef](#)]
22. Jain, S.K. Investigating the behavior of statistical indices for performance assessment of a reservoir. *J. Hydrol.* **2010**, *391*, 90–96. [[CrossRef](#)]
23. Sklar, M. Fonctions de repartition an dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris* **1959**, *8*, 229–231.
24. Nelsen, R.B. *An Introduction to Copulas*; Springer Science & Business Media: Berlin, Germany, 2007.
25. 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. [[CrossRef](#)]
26. Salvadori, G.; De Michele, C. On the use of copulas in hydrology: Theory and practice. *J. Hydrol. Eng.* **2007**, *12*, 369–380. [[CrossRef](#)]
27. Embrechts, P.; Lindskog, F.; McNeil, A. *Modelling Dependence with Copulas and Applications to Risk Management*; Institut Fédéral de Technologie de Zurich: Zurich, Switzerland, 2001.
28. 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. [[CrossRef](#)]
29. Frank, M.J. On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$. *Aequ. Math.* **1979**, *19*, 194–226. [[CrossRef](#)]
30. Hutchinson, T.P. *Continuous Bivariate Distributions Emphasising Applications*; Rumsby Scientific Publishing: Adelaide, Australia, 1990.
31. Schmidt, R.; Stadtmüller, U. Non-parametric estimation of tail dependence. *Scand. J. Stat.* **2006**, *33*, 307–335. [[CrossRef](#)]
32. Zhang, L.; Singh, V.P. Bivariate rainfall frequency distributions using archimedean copulas. *J. Hydrol.* **2007**, *332*, 93–109. [[CrossRef](#)]
33. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723. [[CrossRef](#)]
34. Gringorten, I.I. A plotting rule for extreme probability paper. *J. Geophys. Res.* **1963**, *68*, 813–814. [[CrossRef](#)]
35. Burnham, K.P.; Anderson, D.R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*; Springer Science & Business Media: Berlin, Germany, 2003.
36. Yevjevich, V. *Probability and Statistics in Hydrology*; Water Resources Publications: Littleton, CO, USA, 1972.
37. Rodriguez-Iturbe, I.; Cox, D.R.; Isham, V. Some models for rainfall based on stochastic point processes. *Proc. R. Soc. Lond. A* **1987**, *410*, 269–288. [[CrossRef](#)]
38. Heneker, T.M.; Lambert, M.F.; Kuczera, G. A point rainfall model for risk-based design. *J. Hydrol.* **2001**, *247*, 54–71. [[CrossRef](#)]
39. Morrissey, M.L. Superposition of the neyman–scott rectangular pulses model and the poisson white noise model for the representation of tropical rain rates. *J. Hydrometeorol.* **2009**, *10*, 395–412. [[CrossRef](#)]
40. Kim, D.; Olivera, F.; Cho, H.; Socolofsky, S.A. Regionalization of the modified Bartlett–Lewis rectangular pulse stochastic rainfall model. *Terr. Atmos. Ocean. Sci.* **2013**, *24*, 421–436. [[CrossRef](#)]

41. Teixeira-Gandra, C.F.; Damé, R.d.C.F. Bartlett–Lewis of rectangular pulse modified model: Estimate of parameters for simulation of precipitation in sub-hourly duration. *Engenharia Agrícola* **2014**, *34*, 925–934. [[CrossRef](#)]
42. Ritschel, C.; Ulbrich, U.; Névir, P.; Rust, H.W. Precipitation extremes on multiple timescales–Bartlett–Lewis rectangular pulse model and intensity–duration–frequency curves. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 6501–6517. [[CrossRef](#)]
43. 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. [[CrossRef](#)]
44. Farrar, D.E.; Glauber, R.R. Multicollinearity in regression analysis: The problem revisited. *Rev. Econ. Stat.* **1967**, *49*, 92–107. [[CrossRef](#)]
45. Silvey, S. Multicollinearity and imprecise estimation. *J. R. Stat. Soc. Ser. B Methodol.* **1969**, *31*, 539–552.
46. Cortina, J.M. Interaction, nonlinearity, and multicollinearity: Implications for multiple regression. *J. Manag.* **1993**, *19*, 915–922. [[CrossRef](#)]
47. Wheeler, D.; Tiefelsdorf, M. Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *J. Geogr. Syst.* **2005**, *7*, 161–187. [[CrossRef](#)]
48. Morrow-Howell, N. The M word: Multicollinearity in multiple regression. *Social Work Res.* **1994**, *18*, 247–251. [[CrossRef](#)]
49. Viswesvaran, C. Multiple regression in behavioral research: Explanation and prediction. *Pers. Psychol.* **1998**, *51*, 223–226.
50. Kutner, M.H.; Nachtsheim, C.; Neter, J. *Applied Linear Regression Models*; McGraw-Hill/Irwin: Boston, MA, USA, 2004.
51. Robinson, C.; Schumacker, R.E. Interaction effects: Centering, variance inflation factor, and interpretation issues. *Mult. Linear Regres. Viewp.* **2009**, *35*, 6–11.
52. Marquandt, D. You should standardize the predictor variables in your regression models. Discussion of: A critique of some ridge regression methods. *J. Am. Stat. Assoc.* **1980**, *75*, 87–91. [[CrossRef](#)]
53. Hasenauer, H.; Monserud, R.A. A crown ratio model for austrian Forests. *For. Ecol. Manag.* **1996**, *84*, 49–60. [[CrossRef](#)]
54. Shiu, H.-J. The application of the value added intellectual coefficient to measure corporate performance: Evidence from technological firms. *Int. J. Manag.* **2006**, *23*, 356–365.
55. Akinwande, M.O.; Dikko, H.G.; Samson, A. Variance inflation factor: As a condition for the inclusion of suppressor variable(s) in regression analysis. *Open J. Stat.* **2015**, *5*, 754. [[CrossRef](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).