Estimation of return period and its uncertainty for the recent 2013–2015 drought in the Han River watershed in South Korea

Jin-Young Kim, Byung-Jin So, Hyun-Han Kwon, Tae-Woong Kim and Joo-Heon Lee

ABSTRACT

This study introduces an approach to copula model parameter estimation within a Bayesian framework. The model is applied to estimate the return periods of the 2013–2015 drought in the Han River watershed (South Korea) as observed through a network of 18 rainfall stations in the watershed. For the univariate return periods, some of the 2013–2015 drought durations (i.e., 9–24 months) have return periods of around 30 years while the drought severities (which range from about 400 to 1,400 mm) at certain stations might be regarded as extreme events with return periods in the hundreds of years. The recent drought is extraordinary at many stations in the Han River watershed although the associated uncertainty is quite large (e.g., for the Seoul station, a range of about 180 to 2,250 years). Moreover, it is clearly seen that the unprecedented joint return periods of the 2013–2015 drought at many stations are consistent with the very high drought severity. The joint return period of the drought duration and severity is above 100 years for many of the stations and upwards of 1,000 years for a few stations.

INTRODUCTION

A number of univariate drought frequency methods have been developed and applied to estimate return periods of droughts (Loganathan et al. 1985; Engeland et al. 2004; Arena et al. 2006; Cebrián & Abaurrea 2006; Lana et al. 2008; Sadri & Burn 2012). Extreme drought events are particularly of interest from a hydrologic design perspective. They are best represented considering the multivariate dependence across the duration, severity, and intensity of an event (Favre et al. 2004; Genest et al. 2007; Renard & Lang 2007; Klein et al. 2009).

Modeling the multivariate dependence has been recognized as necessary for hydrologic risk estimation (Shiau 2006; Hao & Singh 2012, 2013; De Michele et al. 2013; Madadgar & Moradkhani 2014; Liu et al. 2015), especially for assessing risks associated with climate change and variability (Lall et al. 2015). For drought risk analyses, hydrologists are interested in how joint return periods or joint likelihoods for drought duration and severity could be estimated to understand their mutual dependences on drought risk (Shiau 2006; Madadgar & Moradkhani 2011; Yoo et al. 2012; Grüler et al. 2013).

Several studies have been conducted to estimate the joint or conditional probability of extreme events with multivariate joint distributions (Raynal-Villasenor & Salas 1987; Yue et al. 2001; Shiau & Shen 2001; Favre et al. 2004; Salas et al. 2005; Shiau 2006; Renard & Lang 2007; Yoo et al. 2012; De Michele et al. 2013; Hao & Singh 2013). A major difficulty in estimating the multivariate joint density function for drought risk analyses is that the marginal distributions of the variables of interest (e.g.,
duration and severity) are often different. Copula-based multivariate frequency techniques have been successfully applied to drought frequency analysis as a solution of the issue of dissimilar marginal distributions (Shiau 2006; Madadgar & Moradkhani 2011; Ma et al. 2013; Zhang et al. 2013; Xu et al. 2013). Copula functions were initially introduced by Sklar (1959) and they provide the flexibility to identify and fit a dependence structure to variables that have different marginal distributions. Despite many successful applications of copula-based frequency analysis in the hydrology community, the issue of estimating the parameter uncertainty is not often addressed. There have been limited studies on the use of Bayesian inference for copula function in hydrology including for estimation of bivariate return periods (Parent et al. 2014; Zhang et al. 2015; Bracken et al. 2016; Kwon & Lall 2016; Kwon et al. 2016; Sarhadi et al. 2016).

Over the last decade, Korea has frequently experienced droughts. Furthermore, future droughts may be more severe due to climate change (Kim et al. 2011; Kwon et al. 2011; Lee et al. 2012; Yoo et al. 2012). Since 2013, South Korea’s water supplies have been significantly stressed as a result of a peninsula-wide ongoing drought. The Han River watershed (i.e., Seoul, Gyeonggi, and Gangwon provinces) has been particularly affected by the recent drought. The region has received about half of its average annual rainfall (1,200 mm). In June 2015, the water level of the Soyang Dam (which is the largest dam in South Korea) was at the lowest level recorded (i.e., 152.25 m a.s.l.) since the dam was built in 1974. In these contexts, this study primarily uses a Bayesian parameter estimation approach to estimate the return periods and their uncertainties associated with the recent 2013–2015 drought events in South Korea.

The data and watershed are introduced immediately below. The theoretical framework for the modeling approach is presented in the section ‘Method and model formulation’. In the following two sections the model is applied to the synthetic data and the case study for the recent 2013–2015 drought in South Korea. The final section presents a summary and conclusions.

HAN RIVER WATERSHED AND DROUGHT DEFINITIONS

We consider the Han River watershed as a case study due to the importance of the watershed from a water security perspective (the watershed is the primary water supply for the Seoul metropolitan area). The total watershed area of the Han River watershed is approximately 35,770 km², and the annual average discharge of the basin is about 613 m³/s. The monthly rainfall data of 18 weather stations for the Han River watershed were obtained from the Korea Meteorological Administration (KMA) (http://web.kma.go.kr/eng), spanning the period 1974–2015. The locations of the gauging stations used in this study are shown in Figure 1.

Meteorological drought can be defined as an extended period of deficient precipitation. In this study, the 6-month rainfall is considered to evaluate seasonal droughts that occur due to rainfall deficit in wet (or monsoon) season and dry season (or non-monsoon). The accumulated 6-month monthly rainfall was transformed to anomaly time series determined by subtracting the long-term mean of the monthly rainfall time series as written in Equation (1). The monthly long-term mean of the full record \( \bar{A}_i \)
was calculated as follows:

\[ A_{(i)} = \frac{R_{1,i} + R_{2,i} + \ldots + R_{n,i}}{n} \]

where \( R_{n,i} \) is the accumulated 6-month rainfall at the end of a month \( i \) and year \( n \).

We considered continuous negative precipitation anomalies to mark a drought, and the first instance of positive precipitation anomalies to mark the end of a drought. In general, the consecutive drought episodes are assumed to be independent realizations. We tested the independence of the extracted consecutive drought episodes (or variables) using autocorrelation function. It was found that the drought variables are linearly independent. As a result of this simple definition, we can define a time series of drought conditions and extract four historic characteristics: drought duration, drought severity, drought intensity (i.e., drought severity/drought duration), and drought interval time. The drought severity is determined as the accumulated rainfall deficit over the corresponding duration.

We extracted drought characteristics from the time series of precipitation anomalies for all the stations in Han River watershed (not all are shown here due to space limitations; please see Figure S1, available with the online version of this paper). Information on the gauging stations used in this study and the basic statistics for the drought characteristics are summarized in Table 1.

We also estimated an areal rainfall by using the Thiessen polygon method and computed watershed-wide historic drought durations and severities (Figures 2 and 3). The drought duration and severity variables are positively skewed (Figure 3), with the recent drought state being the most severe and longest duration drought in the post-1974 record.

**METHODS AND MODEL FORMULATION**

**Copula function overview**

A copula is a multivariate probability function to describe interdependence among multivariate random variables.
The copula has been widely used for multivariate analysis of hydrologic variables because of its flexibility and power to establish a joint probability function with the different marginals. Favre et al. (2004) showed the efficacy of using copula functions in multivariate hydrologic frequency analysis. Next, we briefly discuss the definition of copulas, and in the remaining part of this section, we present the candidate parametric copula models used in this study and discuss Bayesian parameter estimation.

Suppose we have m-dimensional random variables \((X_1, X_2, \ldots, X_m)\) and their marginal cumulative distribution functions (CDFs) (i.e., \(F_i(x) = p[X_i \leq x]\)) are continuous. The random variables can be transformed into uniformly distributed marginal distributions by applying the marginal CDFs to the random variables as in Equation (2):

\[
(U_1, U_2, \ldots, U_m) = (F_1(X_1), F_2(X_2), \ldots, F_m(X_m))
\]

The copula of random variables is then defined as a joint CDF (Equation (3)):

\[
C(u_1, u_2, \ldots, u_m) = p[U_1 \leq u_1, U_2 \leq u_2, \ldots, U_m \leq u_m]
\]

The marginal CDF \(F_i\) describes underlying marginal distribution of variable \(i\), while the copula function \(C\) describes the dependence structure between the random variables \((X_1, X_2, \ldots, X_m)\). A main advantage of the copula approach is that the reverse of the above steps can be applied to efficiently simulate multivariate random samples. Specifically, the required multivariate random variables can be sampled from a uniformly distributed random vector \((U_1, U_2, \ldots, U_m)\) derived from the defined copula function (Equation (4)):

\[
(X_1, X_2, \ldots, X_m) = F_1^{-1}(U_1), F_2^{-1}(U_2), \ldots, F_m^{-1}(U_m)
\]

where \(F_i^{-1}\) is the quantile function (or inverse CDF) of the marginal distribution, and Equation (5) can be expressed...
as Equation (5):

\[
C(u_1, u_2, \ldots, u_m) = p \left[ X_1 \leq F_1^{-1}(u_1), X_2 \leq F_2^{-1}(u_2), \ldots, X_m \leq F_m^{-1}(u_m) \right]
\]  

(5)

**Marginal distribution**

Lognormal (Equation (6)) and Gamma (Equation (7)) distributions were selected as marginal distributions for drought durations (x) and drought severity (y), respectively, based on the Bayesian information criterion (BIC) (see Supplementary information Table S1, available with the online version of this paper). The distribution with the lowest BIC is preferred:

\[
f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-((\log(x) - \mu)^2/2\sigma^2)}
\]  

(6)

\[
f(y) = e^{ky}k^{-1}e^{-y} \Gamma(k)
\]  

(7)

where \(\mu\) is the mean, \(\sigma\) is the standard deviation, \(k\) is the shape parameter, \(r\) is the rate parameter, and \(\Gamma(k)\) is the gamma function evaluated at \(k\).

The scatter plot of the synthetic pairs using a Frank copula with Lognormal and Gamma as the two marginal distributions is shown in Figure 3, for visual representation of the dependence structure. A further discussion on the copula selection is provided in the following section. The empirical CDFs along with the theoretical marginal CDFs for four stations (i.e., 105, 108, 201, and 203) are illustrated in Figure 4.

**Copula types and parameter estimation**

This study employed three types of Archimedean copulas to model dependency structure of drought variables, namely, the Clayton, Frank, and Gumbel. All three of the copula functions are defined by one parameter (\(\theta\)). The Archimedean copula is very flexible and convenient in terms of simulation because a closed functional form is available. The copula parameter, \(\theta\), can also be estimated by using general bivariate dependency measures such as Kendall’s rank correlation coefficient and Spearman’s rank correlation coefficient. The functional forms of the Clayton \((C_C)\), Frank \((C_F)\), and Gumbel \((C_G)\) copulas are given in Equations (8)–(12) as follows:

\[
C_C(u, v; \theta) = \left( \max \left\{ u^{-\theta} + v^{-\theta} - 1; 0 \right\} \right)^{-1/\theta}
\]  

(8)

\[
C_F(u, v; \theta) = -\theta^{-1}\log \left( 1 + \frac{(e^{-u\theta} - 1)(e^{-v\theta} - 1)}{e^{-\theta} - 1} \right)
\]  

(9)

\[
C_G(u, v; \theta) = e^{\left( -\left( -\log(u) \right)^{\theta} + \left( -\log(v) \right)^{\theta} \right)^{1/\theta}}
\]  

(10)
and

\[ u = F_u(x) \quad (11) \]

\[ v = F_v(y) \quad (12) \]

where \( F_u \) and \( F_v \) indicate marginal CDFs for the drought durations (x) and drought severity (y), respectively.

A Bayesian approach to copula and marginal distribution parameter estimation is proposed to better estimate parameter uncertainty. The posterior distribution of the parameter vector, \( \Theta \), \( p(\Theta|D) \) is given by Bayes’ theorem (Equation (13)):

\[ p(\Theta|D) = \frac{p(\Theta) \times p(D|\Theta)}{p(D)} = \frac{p(\Theta) \times p(D|\Theta)}{\int p(\Theta) \times p(D|\Theta)d\Theta} \propto p(\Theta) \times p(D|\Theta) \quad (13) \]

where \( \Theta \) is the parameter vector of the distribution to be fitted, \( p(D|\Theta) \) is the likelihood function, \( D \) is the observation vector, and \( p(\Theta) \) is the prior distribution.

A main advantage of the copula approach is that one can estimate joint distributions in two stages. The marginal distributions can first be fit by assuming that the variables are independent identically distributed vectors, then the dependence between the variables can be incorporated through the copulas. As shown in the previous section, this study assumed that the marginal distributions for drought duration and severity characteristics are Lognormal and Gamma distribution, respectively. Hence, the joint log-likelihood function for the drought characteristics based on the Clayton \( L_C \), Frank \( L_F \), and Gumbel \( L_G \) copula functions along with the Lognormal and the Gamma distributions can be written as Equations (14)–(16):

\[ L_C(x, y; \mu, \sigma, \kappa, \tau, \theta) = -\log(\theta + 1) + (2\theta + 1) \times \left( -\frac{1}{\theta} \right) \]

\[ \times \log(u^\theta + v^\theta - 1) - (\theta + 1) \times (\log(u) + \log(v)) - \log(x) - 0.5 \log(2\tau) - 0.5 \log(\sigma) - 0.5/\sigma \]

\[ \times (\log(x) - \mu)^2 + (k - 1) \times \log(y) - y/\tau - k \log(\tau) - \log(\Gamma(k)) \quad (14) \]

\[ L_F(x, y; \mu, \sigma, \kappa, \tau, \theta) = -\theta(u + v) - 2\log(1 - e^{-\theta} - (1 - e^{-\sigma})(1 - e^{-\kappa})) + \log(\theta) + \log(1 - e^{-\theta}) - \log(x) - 0.5 \log(2\tau) - 0.5 \log(\sigma) - 0.5/\sigma \]

\[ -0.5/\sigma \times (\log(x) - \mu)^2 + (k - 1) \times \log(y) - y/\tau - k \log(\tau) \quad (15) \]

\[ L_G(x, y; \mu, \sigma, \kappa, \tau, \theta) = \log(\theta - 1 + C^\prime) - C^\prime \]

\[ + ((\theta - 1) \times \log(uv_{\text{min}}) + uv_{\text{min}}) + ((\theta - 1) \times \log(uv_{\text{max}}) + uv_{\text{max}}) + (1 - 2\theta) \]

\[ \times \log(C') - \log(x) - 0.5 \log(2\tau) - 0.5 \log(\sigma) - 0.5/\sigma \]

\[ \times (\log(x) - \mu)^2 + (k - 1) \times \log(y) - y/\tau - k \log(\tau) - \log(\Gamma(k)) \quad (16) \]

where, \( x \) and \( y \) indicate drought durations and drought severity, respectively.

The prior distributions for the five parameters (e.g., \( \mu, \sigma, k, \tau, \text{ and } \theta \)) are also assigned. Conjugate distributions are probability distributions whose prior and posterior distributions are in the same family. In particular, conjugate distributions are favored for computational reasons. In these contexts, the mean \( (\mu) \) is assumed to be Gaussian with zero-mean and precision \( 10^{-3} \), the standard deviation, scale, and rate parameters \( (\sigma, k, \text{ and } \tau) \) are assumed to be gamma, and the copula parameter \( (\theta) \) is assumed to be uniform. Here, the prior distributions for the parameters \( \sigma, k, \text{ and } \tau \) were given weak informative gamma priors (i.e., gamma \( (0.1, 0.1) \)), which has mean 1 and variance 10. The Kendal rank correlation coefficients 0.9 is equal to 38.28, 18.00, and 10.00 for Frank, Clayton, and Gumbel copulas, respectively. In this regard, the copula parameter \( \theta \) is given informative uniform prior distributions \( U(0, 50) \) to reflect the expected positive
dependence structure:

μ ~ Normal(0, 0.001) (17)

σ ~ Gamma(0.1, 0.1) (18)

k ~ Gamma(0.1, 0.1) (19)

τ ~ Gamma(0.1, 0.1) (20)

θ ~ Uniform(0, 50) (21)

Combining the prior parameter distributions (as defined in Equations (17)–(21)) and the log-likelihood function (Equations (14)–(16)) into Equation (13) yields the joint log-posterior distribution of the parameters (Equation (22)):

\[
p(\Theta | D) \propto \sum_{i=1}^{n} L(x_i, y_i; \mu, \sigma, \kappa, \tau, \theta) + \text{Normal}(\mu | 0.001) + \text{Gamma}(\sigma | 0.1, 0.1) + \text{Gamma}(k | 0.1, 0.1) + \text{Gamma}(\pi | 0.1, 0.1) + \text{Uniform}(\theta | 0, 50)
\] (22)

The five parameters defined above are fit using the Gibbs sampler, a Markov Chain Monte Carlo (MCMC) method for estimating the posterior probability density function of model parameters (Gelman & Hill 2006). The Gibbs sampler aims to search the multivariate parameter space by:

(1) holding constant all but one of the parameters, (2) simulating from the conditional distribution of the unfixed parameter, and (3) repeating for the rest of the parameters (Kwon et al. 2008). For further information on the use of the Gibbs sampler for Bayesian MCMC, please see Gilks et al. (1995).

**Return periods**

The univariate return period of a certain drought severity (Shiau & Shen 2001) \(T_{DS}\) and drought duration (Shiau 2006) \(T_{DD}\) can be estimated using the formula as follows:

\[
T_{DS} = \frac{E(L)}{P(S \geq s)} = \frac{E(L)}{1 - F_S(s)} \quad (23a)
\]

\[
T_{DD} = \frac{E(L)}{P(D \geq d)} = \frac{E(L)}{1 - F_D(d)} \quad (23b)
\]

where \(E(L)\) is the expected drought interval time, \(s\) and \(d\) denote drought severity and drought duration, and \(F_S\) and \(F_D\) represent cumulative distributions of the drought severity and duration, respectively.

Shiau (2006) proposed a methodology that categorizes the return periods of bivariate distributed hydrologic events as joint and conditional return periods. The joint return period for drought duration and severity can be defined in two cases: return period for \(D \geq d\) and \(S \geq s\) \((T_{DDS})\) and return period for \(D \geq d\) or \(S \geq s\) \((T'_{DDS})\). The joint return period \(T_{DDS}\), the so-called ‘AND’ return period, is that the certain threshold drought duration \((d)\) and severity \((s)\) are exceeded by the respective random variables \(D\) and \(S\) (Requena et al. 2013). On the other hand, the joint return period \(T'_{DDS}\), the so-called ‘OR’ return period is that the certain threshold drought duration \((d)\) or severity \((s)\) are exceeded by the respective random variables \(D\) and \(S\) (Requena et al. 2013):

\[
T_{DDS} = \frac{E(L)}{P(D \geq d \text{ and } S \geq s)} = \frac{E(L)}{1 - F_D(d) - F_S(s) + C(F_D(d), F_S(s))} \quad (24a)
\]

\[
T'_{DDS} = \frac{E(L)}{P(D \geq d \text{ or } S \geq s)} = \frac{E(L)}{1 - C(F_D(d), F_S(s))} \quad (24b)
\]

where \(C(F_D(d), F_S(s))\) represents the joint probability between the drought duration and severity. There is an inequality between the univariate return periods and joint return period (Requena et al. 2013) as follows:

\[
T'_{DDS} \leq \min[T_{DD}, T_{DS}] \leq T_{DDS} \leq T_{DD} \leq T_{DS} \leq T'_{DD} \quad (25)
\]

Detailed discussions on the relationships between the univariate and bivariate return periods can be found in Shiau (2006).

**MODEL VALIDATION WITH SYNTHETIC DATA**

An experimental study was performed by using synthetic data generated from the three copula functions with a set...
of predetermined parameters. We used Lognormal(1, 0.5) and Gamma(3, 1) as the two marginal distributions and used a Spearman’s rank correlation coefficient ($\rho$) of 0.85 to define the dependence between the marginal distributions. Hence, the true copula parameters corresponding to the Spearman’s rank correlation coefficient ($\rho$) of 0.85 for Frank, Clayton, and Gumbel copulas are 4.066, 9.563, and 3.015, respectively, as summarized in Table 2. Five hundred correlated random variables were sampled for each of the three copula functions (Figure 5).

We generated three MCMC chains of length 10,000 and discarded the first 8,000 as burn-in. The estimated posterior densities of the parameters based on 2,000 MCMC samples from three chains are displayed in Table 2 and Figure 6. We then estimated the Bayesian credible region (defined as the interval that contains 95% of the posterior distribution).

The true values of the parameters all fall within the credible region in all experimental cases (Table 2). The biases in the posterior medians compared to the true values are relatively small and may be partially explained by sampling error in the generation of the bivariate samples. Table 2 and Figure 6 suggest that the proposed Bayesian approach is suitable for copula-based multivariate analysis.

Another central question is whether the posterior likelihood can be used to identify the correct generating copula model. To answer this, we again generated three bivariate random samples from the three known copulas. For each of these samples, we fit three models (Clayton, Frank, and Gumbel copulas) and compare the deviance information criterion (DIC) from the posterior likelihood functions of each of the models. The DIC is especially useful for hierarchical Bayesian model selection when the posterior distributions have been derived by MCMC simulation. The DIC is defined as the sum of the posterior mean of the deviance ($D = -2\log L$) and the effective number of parameters ($p_D$) estimated from the posterior distributions of the model parameters.

The estimated posterior distributions of the parameters along with the DIC value are summarized in Table 3. Lower DIC values indicate better model fit to the data. The results indicate that the proposed modeling scheme is capability of correctly identifying the data generating copula function.

The observed drought data used in this work contain ‘Ties’ (i.e., the repeated measurements, especially concerning the duration), which may adversely affect the identification

### Table 2
The experimental study for the three copula functions. The 95% credible intervals for the copula and marginal distribution are estimated from the posterior distributions.

<table>
<thead>
<tr>
<th>Copula</th>
<th>Parameter</th>
<th>True value</th>
<th>2.50%</th>
<th>50%</th>
<th>97.50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clayton copula</td>
<td>$\kappa$</td>
<td>3.000</td>
<td>2.784</td>
<td>2.992</td>
<td>3.215</td>
</tr>
<tr>
<td></td>
<td>$\tau$</td>
<td>1.000</td>
<td>0.912</td>
<td>0.984</td>
<td>1.061</td>
</tr>
<tr>
<td></td>
<td>$\mu$</td>
<td>1.000</td>
<td>0.970</td>
<td>1.003</td>
<td>1.039</td>
</tr>
<tr>
<td></td>
<td>$\sigma$</td>
<td>0.500</td>
<td>0.516</td>
<td>0.497</td>
<td>0.479</td>
</tr>
<tr>
<td></td>
<td>$\theta$</td>
<td>4.066</td>
<td>3.474</td>
<td>3.910</td>
<td>4.394</td>
</tr>
<tr>
<td>Frank copula</td>
<td>$\kappa$</td>
<td>3.000</td>
<td>2.853</td>
<td>3.097</td>
<td>3.361</td>
</tr>
<tr>
<td></td>
<td>$\tau$</td>
<td>1.000</td>
<td>0.941</td>
<td>1.030</td>
<td>1.125</td>
</tr>
<tr>
<td></td>
<td>$\mu$</td>
<td>1.000</td>
<td>0.975</td>
<td>1.006</td>
<td>1.041</td>
</tr>
<tr>
<td></td>
<td>$\sigma$</td>
<td>0.500</td>
<td>0.517</td>
<td>0.496</td>
<td>0.476</td>
</tr>
<tr>
<td></td>
<td>$\theta$</td>
<td>9.563</td>
<td>8.443</td>
<td>9.371</td>
<td>10.290</td>
</tr>
<tr>
<td>Gumbel copula</td>
<td>$\kappa$</td>
<td>3.000</td>
<td>2.613</td>
<td>2.817</td>
<td>3.054</td>
</tr>
<tr>
<td></td>
<td>$\tau$</td>
<td>1.000</td>
<td>0.867</td>
<td>0.940</td>
<td>1.020</td>
</tr>
<tr>
<td></td>
<td>$\mu$</td>
<td>1.000</td>
<td>0.977</td>
<td>1.013</td>
<td>1.046</td>
</tr>
<tr>
<td></td>
<td>$\sigma$</td>
<td>0.500</td>
<td>0.532</td>
<td>0.512</td>
<td>0.492</td>
</tr>
<tr>
<td></td>
<td>$\theta$</td>
<td>3.015</td>
<td>2.777</td>
<td>3.020</td>
<td>3.273</td>
</tr>
</tbody>
</table>

Figure 5 | Synthetic bivariate random variables using (a) Clayton, (b) Frank and (c) Gumbel copula functions.
of the generating copula during the experimental study performed in this study. The identification problem when Ties are present has been investigated and the results were negative (De Michele et al. 2013). In this regard, we performed a number of simulations to test the sensitivity of the Ties’ effects. As performed in the previous experimental study, three bivariate random samples from the three known copulas were generated, and the generated drought duration is then rounded to the nearest integer value to consider the Ties’ effects, as displayed in Figure S2 (available with the online version of this paper). For each of these samples, we fit three models (Clayton, Frank, and Gumbel copulas) and

<table>
<thead>
<tr>
<th>Model</th>
<th>Lognormal</th>
<th>Gamma</th>
<th>Copula</th>
<th>Log-likelihood</th>
<th>DIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case A (Clayton copula)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clayton</td>
<td>2.725</td>
<td>1.155</td>
<td>0.989</td>
<td>0.741</td>
<td>4.066/3.900</td>
</tr>
<tr>
<td>Frank</td>
<td>2.723</td>
<td>1.114</td>
<td>0.984</td>
<td>0.732</td>
<td>9.563/10.663</td>
</tr>
<tr>
<td>Gumbel</td>
<td>2.721</td>
<td>1.161</td>
<td>0.999</td>
<td>0.752</td>
<td>3.015/2.353</td>
</tr>
<tr>
<td>Case B (Frank copula)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clayton</td>
<td>2.757</td>
<td>1.151</td>
<td>0.999</td>
<td>0.747</td>
<td>4.066/2.633</td>
</tr>
<tr>
<td>Frank</td>
<td>2.984</td>
<td>1.048</td>
<td>0.978</td>
<td>0.734</td>
<td>9.563/12.877</td>
</tr>
<tr>
<td>Gumbel</td>
<td>2.817</td>
<td>1.122</td>
<td>0.978</td>
<td>0.758</td>
<td>3.015/3.007</td>
</tr>
<tr>
<td>Case C (Gumbel copula)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clayton</td>
<td>3.472</td>
<td>0.795</td>
<td>0.878</td>
<td>0.662</td>
<td>4.066/2.445</td>
</tr>
<tr>
<td>Frank</td>
<td>3.259</td>
<td>0.865</td>
<td>0.891</td>
<td>0.646</td>
<td>9.563/9.680</td>
</tr>
<tr>
<td>Gumbel</td>
<td>3.457</td>
<td>0.794</td>
<td>0.868</td>
<td>0.612</td>
<td>3.015/3.004</td>
</tr>
</tbody>
</table>

The Log-likelihoods and DIC values in bold are the optimum copulas selected in the experimental study.
compare the DIC from the posterior likelihood functions of each of the models. The estimated posterior distributions of the parameters along with the DIC value are summarized in Table S2 (available online). The results indicate that the proposed modeling scheme is still correctly identifying the data generating copula function. However, for the Clayton copula function, the DIC values are relatively similar to that of the other copulas, meaning that the Clayton copula function is likely sensitive to the Ties.

**BIVARIATE DROUGHT FREQUENCY ANALYSIS**

**Model identification and parameter estimation**

As with the synthetic dataset above, we use the Clayton, Frank, and Gumbel copulas as candidates for modeling dependency structure of drought variables in the Han River watershed. The marginal distributions for drought duration and severity were assumed to be the Lognormal and Gamma, respectively. The DIC value was calculated for the three copulas under the same marginal distributions and used to select the best model. This was done for a network of 18 weather stations (Table 4). The Frank copula is selected for 13 stations, and the Gumbel for the remaining stations. The lower DIC values for the Gumbel copula compared to the Frank copula indicate that upper tail dependence is an important feature of the drought duration and severity data. The selected copula functions are used for subsequent analysis.

The marginal posterior distributions for the parameters for each station are shown in Figure 7. The credible intervals of the marginal posterior distributions are obtained from three chains based on 10,000 iterations and with the first 8,000 discarded as burn-in. The parameters of the marginal distributions associated with drought duration (i.e., μ and σ) appear to be more consistent across stations compared to the parameters of the drought severity (i.e., κ and τ).

**Return period estimation for the 2013–2015 drought**

The univariate drought duration and drought severity associated with the selected return levels (2, 5, 10, 30, 50, and 100-year) are shown in Table S3 (available with the online version of this paper). The expected recurrence interval between successive drought events, E(L), is from about 8 to 12 months across the weather stations. A recommended design criterion for water storage structures including reservoirs is the 30-year return period with either rainfall amounts or discharge in South Korea. The drought durations and severities corresponding to a 30-year return period for weather stations in the Han River watershed have a median (range) of about 23 (20–30) months and 508 (404–965) mm. Considering that many water storage structures are designed for the 30-year drought, a prolonged rainfall deficit of over 500 mm over 20 months could lead to a higher likelihood of sustained water shortage throughout the watershed. Given this, it is startling to note that the rainfall deficits of the 2013–2015 drought range from about 400 to 1,400 mm. The estimated univariate return period for the 2013–2015 drought and severity among the stations analyzed here ranges from about 4 to 31 years and 17 to 6,105 years, respectively. The univariate and joint return periods of the recent 2013–2015 drought are displayed in Table 5. Some of the 2013–2015 drought durations

**Table 4 | DIC values of the three different copulas for the weather stations**

<table>
<thead>
<tr>
<th>Station no.</th>
<th>Station name</th>
<th>Clayton</th>
<th>Frank</th>
<th>Gumbel</th>
</tr>
</thead>
<tbody>
<tr>
<td>90</td>
<td>Sokcho</td>
<td>698.897</td>
<td>670.458</td>
<td>680.434</td>
</tr>
<tr>
<td>100</td>
<td>Daeqwaleong</td>
<td>664.055</td>
<td>650.905</td>
<td>650.537</td>
</tr>
<tr>
<td>101</td>
<td>Chuncheon</td>
<td>845.887</td>
<td>819.694</td>
<td>828.498</td>
</tr>
<tr>
<td>105</td>
<td>Gangneung</td>
<td>759.797</td>
<td>716.386</td>
<td>708.380</td>
</tr>
<tr>
<td>108</td>
<td>Seoul</td>
<td>860.898</td>
<td>844.142</td>
<td>846.563</td>
</tr>
<tr>
<td>112</td>
<td>Incheon</td>
<td>837.784</td>
<td>825.275</td>
<td>828.659</td>
</tr>
<tr>
<td>114</td>
<td>Wonju</td>
<td>837.889</td>
<td>820.206</td>
<td>830.867</td>
</tr>
<tr>
<td>119</td>
<td>Suwon</td>
<td>799.869</td>
<td>793.611</td>
<td>792.983</td>
</tr>
<tr>
<td>131</td>
<td>Chongju</td>
<td>811.919</td>
<td>785.285</td>
<td>782.707</td>
</tr>
<tr>
<td>201</td>
<td>Ganghwa</td>
<td>890.634</td>
<td>862.785</td>
<td>881.893</td>
</tr>
<tr>
<td>202</td>
<td>Yangpyeong</td>
<td>859.953</td>
<td>844.063</td>
<td>845.178</td>
</tr>
<tr>
<td>203</td>
<td>Icheon</td>
<td>844.625</td>
<td>832.102</td>
<td>842.248</td>
</tr>
<tr>
<td>211</td>
<td>Inje</td>
<td>803.735</td>
<td>789.577</td>
<td>789.920</td>
</tr>
<tr>
<td>212</td>
<td>Hongcheon</td>
<td>813.667</td>
<td>794.247</td>
<td>801.213</td>
</tr>
<tr>
<td>221</td>
<td>Jecheon</td>
<td>825.814</td>
<td>821.015</td>
<td>813.988</td>
</tr>
<tr>
<td>226</td>
<td>Boeun</td>
<td>745.422</td>
<td>721.203</td>
<td>724.536</td>
</tr>
<tr>
<td>272</td>
<td>Youngju</td>
<td>820.933</td>
<td>801.918</td>
<td>805.088</td>
</tr>
<tr>
<td>273</td>
<td>Mungyeong</td>
<td>714.144</td>
<td>696.728</td>
<td>698.273</td>
</tr>
</tbody>
</table>

The DIC values in bold are the optimum copulas selected in this study.
Figure 7 | Boxplots of the marginal posterior distributions for the parameters obtained from three chains based on 10,000 iterations. The symbol ‘-' indicates a median value of the posteriors. The stations, labeled as 100, 105, 119, 131 and 221, are based on the Gumbel copula while Frank copula for the remains of stations.

Table 5 | The univariate and joint median return periods and their 95% credible intervals for the 2013–2015 drought event

<table>
<thead>
<tr>
<th>Station no.</th>
<th>$T_{DD}$</th>
<th>$T_{DS}$</th>
<th>$T_{DDS}$</th>
<th>$T_{DSS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>90</td>
<td>12.35 (8.58, 18.73)</td>
<td>62.10 (36.40, 142.14)</td>
<td>85.07 (36.90, 228.67)</td>
<td>11.70 (8.31, 17.48)</td>
</tr>
<tr>
<td>100</td>
<td>12.74 (8.69, 19.69)</td>
<td>25.92 (15.59, 47.27)</td>
<td>26.17 (15.76, 47.69)</td>
<td>12.69 (8.65, 19.57)</td>
</tr>
<tr>
<td>101</td>
<td>25.22 (15.24, 43.64)</td>
<td>1,484.67 (392.77, 7,366.58)</td>
<td>4,601.68 (859.00, 30,572.91)</td>
<td>24.90 (15.07, 45.16)</td>
</tr>
<tr>
<td>105</td>
<td>14.25 (9.59, 21.70)</td>
<td>17.19 (11.11, 27.94)</td>
<td>18.92 (12.24, 30.49)</td>
<td>13.20 (8.97, 19.97)</td>
</tr>
<tr>
<td>108</td>
<td>31.07 (18.27, 57.05)</td>
<td>554.38 (180.16, 2,253.62)</td>
<td>2,048.99 (472.31, 11,989.83)</td>
<td>29.72 (17.62, 54.57)</td>
</tr>
<tr>
<td>112</td>
<td>24.65 (15.47, 41.56)</td>
<td>398.42 (131.63, 1,252.64)</td>
<td>1,125.54 (275.09, 4,812.24)</td>
<td>23.60 (14.83, 39.83)</td>
</tr>
<tr>
<td>114</td>
<td>4.86 (3.85, 6.27)</td>
<td>19.38 (12.17, 33.50)</td>
<td>23.31 (13.84, 42.44)</td>
<td>4.67 (3.75, 5.90)</td>
</tr>
<tr>
<td>119</td>
<td>30.03 (17.74, 54.81)</td>
<td>83.38 (39.65, 196.47)</td>
<td>84.35 (40.37, 197.69)</td>
<td>29.88 (17.67, 54.44)</td>
</tr>
<tr>
<td>131</td>
<td>4.50 (3.61, 5.80)</td>
<td>22.43 (13.89, 38.48)</td>
<td>22.53 (13.98, 38.67)</td>
<td>4.49 (3.60, 5.79)</td>
</tr>
<tr>
<td>201</td>
<td>29.68 (17.16, 55.15)</td>
<td>6,104.60 (1,165.77, 43,064.00)</td>
<td>22,264.77 (2,977.98, 227,804.71)</td>
<td>29.54 (17.10, 57.49)</td>
</tr>
<tr>
<td>202</td>
<td>23.75 (14.71, 41.42)</td>
<td>393.50 (133.86, 1,524.36)</td>
<td>1,216.76 (309.56, 6,712.96)</td>
<td>22.65 (14.11, 39.29)</td>
</tr>
<tr>
<td>203</td>
<td>4.88 (3.88, 6.38)</td>
<td>33.37 (19.29, 63.64)</td>
<td>39.88 (22.21, 80.03)</td>
<td>4.77 (3.82, 6.16)</td>
</tr>
<tr>
<td>211</td>
<td>22.26 (13.82, 59.16)</td>
<td>176.96 (68.42, 552.63)</td>
<td>459.47 (129.50, 1,989.27)</td>
<td>20.62 (12.97, 35.62)</td>
</tr>
<tr>
<td>212</td>
<td>22.01 (13.82, 59.16)</td>
<td>1,059.75 (272.35, 4,667.83)</td>
<td>2,765.07 (538.07, 16,453.46)</td>
<td>27.71 (13.63, 37.43)</td>
</tr>
<tr>
<td>221</td>
<td>5.25 (4.04, 6.95)</td>
<td>17.75 (11.21, 29.17)</td>
<td>18.18 (11.46, 29.92)</td>
<td>5.21 (4.01, 6.88)</td>
</tr>
<tr>
<td>226</td>
<td>4.22 (3.46, 5.29)</td>
<td>17.32 (10.84, 30.17)</td>
<td>18.33 (11.30, 32.56)</td>
<td>4.17 (3.44, 5.19)</td>
</tr>
<tr>
<td>272</td>
<td>4.56 (3.68, 5.85)</td>
<td>24.12 (14.22, 44.90)</td>
<td>27.28 (15.47, 52.86)</td>
<td>4.47 (3.62, 5.66)</td>
</tr>
<tr>
<td>273</td>
<td>4.38 (3.56, 5.56)</td>
<td>20.24 (12.26, 36.62)</td>
<td>21.94 (12.95, 41.23)</td>
<td>4.30 (3.53, 5.41)</td>
</tr>
</tbody>
</table>

The values in parentheses show the 95% credible intervals.
(i.e., 9–24 months) have median return periods of around 30 years while the drought severities at certain stations may be regarded as extreme events with return periods in the hundreds of years. It is clear that the recent drought is extraordinary at many stations in the Han River watershed although the associated uncertainty is quite large (the 95% credible interval at the Seoul station, for example, has a range of about 180 to 2,250 years). Given these results and the inequality described in Equation (25), we see that the joint return periods are much higher than those of the univariate.

The joint distribution of drought duration and severity based on the copula functions and the Lognormal and Gamma marginal distributions was used to estimate the joint return periods for $T_{DDS}$ (duration and severity, $D \geq d$ and $S \geq s$) and $T_{DDS}$ (duration or severity, $D \geq d$ or $S \geq s$) using Equations 24(a) and 24(b) (Table 5). There was a range of 40 to 59 drought events (as per our precipitation anomaly-based definition in the section ‘Han River watershed and drought definitions’) over the 42 years across the weather stations. The joint return period ($T_{DDS}$) for the 2013–2015 drought is, in general, estimated to be greater than 100 years; estimated return periods are over 1,000 years at some stations. It is clear that the unprecedented joint return periods of the 2013–2015 drought at many stations are consistent with the very high drought severity of the current event. Contour plots of the joint return periods along with the 95% Bayesian credible intervals for selected stations are shown in Figure 8. The figure shows that the uncertainty ranges increase as the return periods increase. The contours also illustrate the variety of combinations of drought durations and severities associated with a given joint return period. The most severe state for the 2013–2015 drought (as measured by the joint return period) is associated with a drought severity of about 1,400 mm and a drought duration of 24 months at station 201 (Ganghwa).

A spatial distribution of the estimated median joint return period of the 2013–2015 drought for the Han River watershed (along with the Bayesian 95% credible interval (Figure S3, available online)) is illustrated in Figure 9. Here, a biharmonic spline interpolation technique is used to construct the map of return period. The western part of the Han River watershed has, in general, experienced the most severe droughts in recorded history over the last four decades.

Finally, the conditional return periods of the drought duration and severity, given drought severity and duration exceeding a certain threshold, were estimated for the four stations, respectively, as shown in Figure 10. Please refer to Shiau (2006) for more detailed formulation of the conditional return periods. Figure 10(a)–10(d) demonstrate the conditional return periods of the drought duration given the drought severity while Figure 10(e)–10(h) represent the conditional return periods of the drought severity given the drought duration. For example, the estimated conditional return period of drought severity given a duration of 24 months for the station 108 (Seoul) is over 100 years.

**SUMMARY AND CONCLUSIONS**

A copula-based multivariate frequency analysis was employed to model the dependences in drought duration and severity. Copula-based multivariate models have been widely applied to hydrologic frequency analyses, however, reliable estimation of the parameters is often a challenging task in the multivariate setting. Moreover, uncertainties associated with the model parameters in copula-based frequency models are not often properly addressed. We have used a Bayesian approach for parameter estimation in a copula-based bivariate drought frequency analysis. The proposed model was applied to estimate joint return periods on a network of 18 rainfall stations in the Han River watershed in South Korea. The key results and conclusions are summarized as follows:

1. A validation of the proposed Bayesian parameter estimation approach was conducted based on synthetic data that were generated by a known set of parameters along with the different copulas (Clayton, Frank, and Gumbel). The validation exercise showed that the true underlying parameters could be reasonably estimated.

2. For the univariate return periods, some of the 2013–2015 drought durations (i.e., 9–24 months) had return periods exceeding certain thresholds at some stations. It was clear that the recent drought was extraordinary at many stations in the Han River watershed although the associated uncertainty is
Figure 8 | The joint return periods of droughts based on Bayesian Copula model for the four stations (i.e. 105, 108, 201 and 203). The symbol ‘filled star’ indicates the 2013–2015 drought and the shaded area indicates the uncertainty associated with the return periods.
quite large (e.g., for the Seoul station, a range of about 180 to 2,250 years).

3. The main point of the paper was to show that the 2013–2015 drought is very rare. We confirmed this finding by doing the uncertainty analysis in the Bayesian framework. It was clear that the unprecedented univariate and joint return periods of the 2013–2015 drought at many stations were consistent with the very high drought

Figure 9 | Spatial distribution of the posterior median of the joint return periods. The symbol ’_’' indicates the stations used in this study. The thick black line encloses sub-catchment region.
Figure 10 | The left panels show the conditional return periods of drought duration given the drought severity while the right panels are the condition return periods of drought severity given the drought duration for the stations 105, 108, 201 and 203. The symbol ‘filled star’ represents the 2013-2015 drought state.
severity of the current event. The estimated joint return period of the drought duration and severity were greater than 100 years for many stations and upwards of 1,000 years for a few stations. The most severe state for the 2015–2015 drought was associated with a drought severity of about 1,400 mm and a drought duration of 24 months at the station 201 (Ganghwa).

However, it should be noted that the estimated return periods for some stations are rather unrealistic and their uncertainties are quite large. This is mainly associated with a relatively small sample size, which introduces sampling uncertainty and variation about the parameters, in addition to the usual level of uncertainty. We have not intended to explore the potential reduction of uncertainties in estimating the return periods at this moment and future work will focus on the use of a hierarchical Bayesian model within a regional drought frequency analysis.

ACKNOWLEDGEMENTS

We thank Upmanu Lall and David Farnham from the Water Center, Columbia University, for providing feedback on the manuscript. We would like to thank the two anonymous reviewers for providing us with suggestions and insightful comments that improved the manuscript. This research was supported by a Grant (15AWMP-B079625-02) from Water Management Research Program funded by Ministry of Land, Infrastructure and Transport of the Korean Government. The data used in this study are available upon request from the corresponding author via email (hkwon@jbnu.ac.kr). The authors declare that there is no conflict of interest regarding the publication of this article.

REFERENCES


Kwon, H. H., Lall, U. & Kim, S. J. 2016 The unusual 2013–2015 drought in South Korea in the context of a multicentury precipitation record: inferences from a nonstationary,


Sklar, M. 1959 *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8, Paris, France.


