A Bivariate Frequency Analysis of Extreme Wave Heights and Periods Using a Copula Function in South Korea

Authors: Kim, Yong-Tak, Park, Jong-Hyeon, Choi, Byoung-Han, Kim, Dong Ho, and Kwon, Hyun-Han

Source: Journal of Coastal Research, 85(sp1) : 566-570

Published By: Coastal Education and Research Foundation

URL: https://doi.org/10.2112/SI85-114.1
A Bivariate Frequency Analysis of Extreme Wave Heights and Periods Using a Copula Function in South Korea

Yong-Tak Kim†, Jong-Hyeon Park†, Byoung-Han Choi††, Dong Ho Kim†††, and Hyun-Han Kwon†††

†Department of Civil Engineering, Chonbuk National University, Korea
††Rural Research Institute, Korea Rural Community Corporation
†††Operational Oceanography Research Center, KIOST, Busan, Korea

ABSTRACT


Sea storms are generally described by a set of variables such as wave height, wave period and wave direction, and these random variables are typically treated as independent of one another. Moreover, univariate wave frequency analysis has been applied to estimate design wave heights corresponding to various return levels (e.g., 20, 50 and 200-year) using significant wave heights under stationary conditions. However, it has been acknowledged that these variables are often correlated with each other, and such dependence structure needs to be considered in the estimation of extreme quantiles. More specifically, a joint estimation of quantiles for different combinations of these variables such as wave heights and wave periods is required to reliably assess optimal design of coastal (or offshore) structures. Over the last several decades, accelerated sea level rise (SLR) and its impact on coastal areas have been reported in many parts of the world. Estimation of extreme quantiles of SLR as a nonstationary process play a crucial role in assessing these impacts. In these contexts, a multivariate frequency model using a copula function approach is introduced to describe sea storm risk, which is mainly characterized by wave heights and periods. The proposed multivariate frequency analysis offers several advantages over widely used univariate stationary frequency analysis including uncertainty estimation, improved representation of inter-dependency and significant improvement of compound risk estimation.

ADDITIONAL INDEX WORDS: Copula, Multivariate frequency analysis, Sea level rise, Wave height, Non-stationarity.

INTRODUCTION

Global warming and climate change have made a large impact on the entire earth system including air, sea, glaciers and land. Most of the impact is expected to be negative (Cho and Maeng, 2007). Recently, as global warming and changes in the frequency of typhoons, resulting storm surges and the high waves cause damage to coastal areas, increasing the risk of damage across all coasts around the globe.

In general, South Korea suffers from storm surges due to typhoons and extratropical cyclones, which are caused by the tide, changes in long-term sea level due to storm surges and resulting high waves. If a storm surge is overlaid with the high tide, sea water invades the coastline over the marine structure, leading to significant damage to property, and people living in the coastal area. For an accurate prediction of a storm surge, it is necessary to consider changes in the tide and waves (Prandle and Worf, 1978; Mastenbroek et al., 1993). In general, the tide is considered in the design practices of marine structures along the coastline, which can manage changes in the tide. However, it is important to analyze waves accurately for prediction of storm surges.

In the energy exchange between the air and sea, the wave is the critical medium (Hemer et al., 2012). There are many studies that have evaluated vulnerability at sea and coastal areas. These studies help people to understand the frequency of waves within climate change (Sterl et al., 1998; Cox and Swail 2001; Wang and Swail 2002). Wang and Swail (2001) and Sasaki et al. (2005) showed that there has been an increase in significant wave height in the winter season in the Northern Pacific region since 1960 and an increase in the period of wave height along the southern coast of Japan due to strengthening typhoons during the same period. Yong et al. (2008) showed that due to the weak summer monsoon in Asia, extreme waves decreased, but due to the strengthening of typhoons occurring in the North Pacific, extreme waves increased in the southern area of the East China Sea.

Korea’s Ministry of Ocean and Fisheries (MOF) analyzed the seasonal winds using the typhoons that occurred during 1951 to 2003 and the seasonal winds that occurred during 1979 to 2003, to better characterize the information on wave design on the three side seas of Korea (Korea Ocean Research & Development Institute II, 2005). These data have been used for the design of coastal structures in Korea up to now. As the Korean peninsula is surrounded by seas on three sides, an increase in wave heights is likely to be critical in coastal areas. Therefore, this study is designed to quantitatively evaluate the waves in South Korea.

In this study, wave frequency was evaluated using the Copula function based bivariate frequency analysis technique as it can evaluate wave height and wave period at the same time for wave evaluation. This paper is composed as follows. First, the data provided by Korea Meteorological Administration (KMA) and

DOI: 10.2112/SI85-114.1 received 30 November 2017; accepted in revision 10 February 2018.
* Corresponding author: hkwon@jbnu.ac.kr
†Coastal Education and Research Foundation, Inc. 2018
European Centre for Medium-Range Weather Forecasts (ECMWF) are compared and analyzed to construct a dataset encompassing more than 30 years. Second, the analysis procedure for the results from the view of statistical inference is provided. Lastly, the results of frequency analysis of the waves are provided along with some suggestions for future research directions.

### METHODS

**Copula**

Copula is a function that represents a multivariate distribution, which considers the association or dependence between multivariate variables. In general, to derive the joint probability density function of two or more variables, it is generally required to interpret the assumption that the marginal probability density function of each variable follows the same distribution. For this purpose, there are many cases in which two-variable Gamma distributions (Yue, 2001) or two-variable Gumbel distributions (Kwon et al., 2016a; Kwon et al., 2016b; Lee et al., 2009) are used for hydrologic frequency analysis.

However, in many cases the marginal distribution of the variables is different. In this case, it is necessary to convert the variables to apply the existing multivariate distribution. A copula function has been proposed to solve this problem. Correlation coefficients are commonly used in multivariate analysis. However, when evaluating the dependency of the tail of the frequency analysis, it is advantageous to use the copula function to grasp its dependency. Sklar’s theorem, named after Abe Sklar, provides the theoretical foundation for the application of copulas. Sklar’s theorem states that every multivariate cumulative distribution function (Equation [1]) of a random vector \( X_1, X_2, ..., X_n \) can be expressed in terms of its marginal cumulative distribution functions (CDFs) (i.e., \( F_i(x) = p[X_i \leq x] \)) and a copula \( C \) (Durante et al., 2000).

\[
C(x_1, x_2, x_3, ..., x_n) = p[X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n] \tag{1}
\]

Suppose we have m-dimensional random variables \( (X_1, X_2, ..., X_m) \) and their marginal CDFs 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).

\[
(U_1, U_2, ..., U_m) = (F_1(X_1), F_2(X_2), ..., F_m(X_m)) \tag{2}
\]

The copula of \( (X_1, X_2, ..., X_m) \) is defined as the joint CDF of \( (U_1, U_2, ..., U_m) \) (Equation (3)).

\[
C(u_1, u_2, u_3, ..., u_m) = p[U_1 \leq u_1, U_2 \leq u_2, ..., U_m \leq u_m] \tag{3}
\]

The marginal CDF \( (F_i) \) contains all information on the marginal distributions, whereas the copula contains all information on the dependence structure between the components of \( (X_1, X_2, ..., X_m) \).

A main advantage of the copula approach is that the reverse of the above steps can be applied efficiently to simulate multivariate random samples. Specifically, the required multivariate random variables can be sampled from a uniformly distributed random vector \( (U_1, U_2, ..., U_m) \) derived from the defined copula function (Equation (4)).

\[
(X_1, X_2, ..., X_m) = (F_1^{-1}(U_1), F_2^{-1}(U_2), ..., F_m^{-1}(U_m)) \tag{4}
\]

Where, \( F_i^{-1} \) is the quantile function (or inverse CDF) of the marginal distribution. The \( F_i^{-1} \) are unproblematic as the \( F_i \) were assumed to be continuous.

\[
C(u_1, u_2, ..., u_m) = p[X_1 \leq F_1^{-1}(u_1), X_2 \leq F_2^{-1}(u_2), ..., X_m \leq F_m^{-1}(u_m)] \tag{5}
\]

Two well known, common families of copulas are the elliptical and Archimedean copulas.

The elliptical copula has two functions, which are the normal copula and Student-\( t \) copula. Simulation of elliptical distribution is easy by linear transformations from standard elliptical distributions. Additionally, one of the main advantages of elliptical copulas is that they can specify different levels of correlation between the margins. For elliptical copulas, there needs to be an estimate of the relationship between the correlation coefficient \( \rho \) and Kendall’s tau \( \tau \) as given by Equation (6).

\[
\rho(x, y) = \sin(\frac{\pi}{2} \tau) \tag{6}
\]

The \( d \)-dimensional normal copula has the following expression (Equation (7)).

\[
C(u_1, u_2, ..., u_d) = \phi_R(\phi_1^{-1}(u_1), \phi_2^{-1}(u_2), ..., \phi_d^{-1}(u_d)) \tag{7}
\]

Where, \( \phi_R \) denotes the \( d \)-dimension standard normal cumulative distribution and \( R \) denotes the corresponding correlation matrix. The density can be written as Equation (8).

\[
C(u_1, u_2, ..., u_d) = \frac{1}{\sqrt{\det(I-\rho \cdot \Sigma)}} \exp(-\frac{1}{2}(\phi_1^{-1}(u_1))^T (\Sigma^{-1} - I) \cdot (\phi_1^{-1}(u_1))) \tag{8}
\]

Where, \( I \) denotes the \( d \times d \) identity matrix.

The Student-\( t \) copula of elliptical copula has the following expression (Equation (9)).

\[
C(u_1, u_2, ..., u_d) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{t_k(v)}(x)d\alpha d\beta d\gamma \tag{9}
\]

Where, \( f_{t_k(v)} \) denotes the \( d \)-dimensional Student-\( t \) density function with degree of freedom \( v \) and \( t_k^{-1} \) denotes the quantile function of a standard univariate Student-\( t \) distribution with degree of freedom \( v \).

On the other hand, the Archimedean copula has been widely applied to bivariate frequency analysis. The Archimedean copula was recognized by Schweizer and Sklar (1961) and Ling (1965). Although elliptical copulas can be easily applied, they do not have explicit expressions and are restricted to the property of radial symmetry. The Archimedean copulas allow for a large variety of different dependence structures and also have explicit expressions.
A Nonstationary Bivariate Frequency Analysis of Extreme Wave Heights and Periods Using a Dynamic Copula Function

Compared with elliptical copulas, the Archimedean copulas are not derived from multivariate distributions using Sklar’s Theorem. Five kinds of copula functions (Gaussian copula, Clayton copula, Frank copula, Gumbel copula and Student t-copula) were considered in the bivariate frequency analysis (Table 1). The likelihood function was estimated and used to select the optimal copula function for each wave station. Archimedean copulas are widely used in hydrology because they allow modeling dependence in high dimensions with only one parameter governing the strength of dependence.

Table 1. Copula functions for bivariate wave height frequency analysis

<table>
<thead>
<tr>
<th>Copula</th>
<th>Bivariate Copula (C_\theta(a, b))</th>
<th>Parameter (\theta)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clayton</td>
<td>([\max(a^{-\theta} + b^{-\theta} - 1; 0)]^{-1/\theta})</td>
<td>(0 \in (-1, \infty))</td>
</tr>
<tr>
<td>Frank</td>
<td>(\frac{1}{\theta} \log (1 + \frac{(\exp(-0a) - 1)(\exp(-0b) - 1)}{\exp(-0) - 1}))</td>
<td>(0 \in (-1, \infty))</td>
</tr>
<tr>
<td>Gumbel</td>
<td>(\exp[-((\log(a))^{\theta} + (\log(b))^{\theta})^{1/\theta}])</td>
<td>(# \in [1, \infty])</td>
</tr>
<tr>
<td>Gaussian</td>
<td>(\frac{\exp(-\frac{\Phi_2^{-1}(u_2)}{\phi_2^{-1}(u_2)})}{\exp(-\frac{\Phi_2^{-1}(u_1)}{\phi_2^{-1}(u_1)})} \times (\sum_i^{-1} - i) \times \frac{\Phi_2^{-1}(u_i)}{\phi_2^{-1}(u_i)})</td>
<td>(# \in [1, \infty])</td>
</tr>
<tr>
<td>Student-t</td>
<td>(1 - [(1 - u)^{\theta} + (1 - v)^{\theta} - (1 - u)^{\theta}(1 - v)^{\theta})^{1/\theta})</td>
<td>(# \in [1, \infty])</td>
</tr>
</tbody>
</table>

In the copula functions, \(u\) and \(v\) are the CDF of their random variables and are parameters of the copula. The \(u\) and \(v\) have the range of 0 to 1, while \(\theta\) has the range written in Table 1.

RESULTS

ERA-Interim Data

For reliable frequency analysis, long-term wave data for at least 20 years are required. Although the observation of wave heights has been measured by KMA and Korea Hydrographic and Oceanographic Agency (KHOA), a reliable frequency analysis is still difficult due to the short length of the data. Accordingly, this study used the ERA-Interim reanalysis data (https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim) for 38 years (from 1979 to the present) provided by European Centre for Medium-Range Weather Forecasts (ECMWF). The annual maximum wave heights and their corresponding wave periods were first extracted. As the ERA-Interim data is the modelled products, we compared the ERA-Interim wave data with the in-situ data from the marine observation buoys for 17 locations, as shown in Figure 1. The correlation between data provided by KMA and ERA-Interim was about 0.78 on average. The results are summarized in Figure 2 and Table 2. Among 17 stations, four representative stations-Ulleungdo, Marado, Donghae, Buan are mainly illustrated in Figure 2.

Bivariate Frequency Analysis

For the bivariate frequency analysis using Copula functions, the first step is to select the marginal probability functions for wave duration and height. Then the Copula function needs to be determined to consider the dependence structure among the wave characteristics.

In order to determine the optimal marginal distribution function, the log-likelihood function was estimated for the various continuous probability distributions. Among possible distributions, the Gumbel distribution and Gamma distribution were selected for the wave height and duration, respectively. Once the marginal probability distribution was determined, the Copula function for each station was then evaluated by log-likelihood function, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). In this study, five different Copula functions were considered for the dependence structure among the wave characteristics. The Gaussian copula was found
to be best for six stations including Chlibaldo, Geomundo and Chujado. The Frank copula was found to be best for eight stations including Ulleungdo, Deokjeokdo and Pohang while the Gumbel and Clayton copulas were selected for the remaining three stations.

Figure 3: Bivariate return periods for wave height and period for four representative stations (stars indicate the recent extreme events) - (a: Ulleungdo, b: Donghae, c: Marado, d: Buan)

Table 2. Selected copula functions for each station

<table>
<thead>
<tr>
<th>Station Name</th>
<th>Location</th>
<th>Correlation</th>
<th>Selected Copula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ulleungdo</td>
<td>37.46</td>
<td>131.11</td>
<td>0.88 Frank</td>
</tr>
<tr>
<td>Deokjeokdo</td>
<td>37.24</td>
<td>126.02</td>
<td>0.81 Frank</td>
</tr>
<tr>
<td>Chlibaldo</td>
<td>34.79</td>
<td>125.78</td>
<td>0.83 Gaussian</td>
</tr>
<tr>
<td>Geomundo</td>
<td>34.00</td>
<td>127.50</td>
<td>0.61 Gaussian</td>
</tr>
<tr>
<td>Geojeo</td>
<td>34.77</td>
<td>128.90</td>
<td>0.66 Clayton</td>
</tr>
<tr>
<td>Donghae</td>
<td>37.54</td>
<td>130.00</td>
<td>0.81 Gumbel</td>
</tr>
<tr>
<td>Pohang</td>
<td>36.35</td>
<td>129.78</td>
<td>0.85 Frank</td>
</tr>
<tr>
<td>Marado</td>
<td>33.08</td>
<td>126.03</td>
<td>0.87 Frank</td>
</tr>
<tr>
<td>Oeyeondo</td>
<td>36.25</td>
<td>125.75</td>
<td>0.87 Frank</td>
</tr>
<tr>
<td>Shiman</td>
<td>34.73</td>
<td>126.24</td>
<td>0.82 Frank</td>
</tr>
<tr>
<td>Chujado</td>
<td>33.79</td>
<td>126.14</td>
<td>0.84 Gaussian</td>
</tr>
<tr>
<td>Incheon</td>
<td>37.09</td>
<td>125.43</td>
<td>0.79 Gaussian</td>
</tr>
<tr>
<td>Buan</td>
<td>35.66</td>
<td>125.81</td>
<td>0.79 Frank</td>
</tr>
<tr>
<td>Seogwipo</td>
<td>33.13</td>
<td>127.02</td>
<td>0.60 Gaussian</td>
</tr>
<tr>
<td>Tongyeong</td>
<td>34.39</td>
<td>128.23</td>
<td>0.63 Frank</td>
</tr>
<tr>
<td>Ulsan</td>
<td>35.35</td>
<td>129.84</td>
<td>0.74 Clayton</td>
</tr>
<tr>
<td>Uljin</td>
<td>36.91</td>
<td>129.87</td>
<td>0.77 Gaussian</td>
</tr>
</tbody>
</table>

Figure 4. Joint Return Periods between wave height and period for the year-2016, for all wave stations in Korea.

CONCLUSIONS

Typhoons have occurred in coastal villages over the Korean Peninsula, causing vast damages to the properties and people. We developed a bivariate frequency analysis model for wave height and period using copula functions. For selection of the optimal

Journal of Coastal Research, Special Issue No. 85, 2018
A Nonstationary Bivariate Frequency Analysis of Extreme Wave Heights and Periods Using a Dynamic Copula Function

copula function for each station, the log-likelihood function was mainly used. Frank and Gaussian copulas were found to be the most suitable for the bivariate frequency analysis. The bivariate frequency analysis using copula functions showed that the return periods for the Southern coast near Jejudo and the Eastern coast near Ulleungdo, which are in the main tracks of typhoons over the Korean Peninsula, were relatively high. It is expected that the results of the bivariate frequency analysis in this study will provide preliminary analysis for the risk analysis associated with wave heights. For future study, a Bayesian copula model will be studied to better quantify the uncertainty of the return periods.

ACKNOWLEDGMENTS

This work is supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant 17AWMP-B083066-04).

LITERATURE CITED


