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

Journal of Coastal Research	SI	85	566–570	Coconut Creek, Florida	2018

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

www.cerf-jcr.org

www.JCRonline.org

ABSTRACT

Kim, Y.-T.; Park, J.-H.; Choi, B.-H.; Kim, D.H., and Kwon, H.-H., 2018. A Bivariate Frequency Analysis of Extreme Wave Heights and Periods Using a Copula Function in South Korea. *In:* Shim, J.-S.; Chun, I., and Lim, H.S. (eds.), *Proceedings from the International Coastal Symposium (ICS) 2018* (Busan, Republic of Korea). *Journal of Coastal Research*, Special Issue No. 85, pp. 566–570. Coconut Creek (Florida), ISSN 0749-0208.

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

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 distribution function like the frequency analysis, it is advantageous to use the copula function to grasp its dependency structure. (Sklar, 1959).

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 $X_1, X_2, \ldots, X_n$ can be expressed in terms of its marginal cumulative distribution functions (CDFs) (*i.e.* $F_i(x) = p[X_i \le x]$) and a copula C (Durante *et al.*, 2000).

$$\mathrm{C}(x_1, x_2, x_3, \ldots, x_n) = P[X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n] \quad (1)$$

Suppose we have m-dimensional random variables $(X_1, X_2, \ldots, X_m)$ 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).

$$(U_1, U_2, \ldots, U_m) = (F_1(X_1), F_2(X_2), \ldots, F_m(X_m)) \quad (2)$$

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

$$\mathrm{C}(u_1, u_2, u_3, \ldots, u_m) = p[U_1 \le u_1, U_2 \le u_2, \ldots, U_m \le u_m] \quad (3)$$

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 $(X_1, X_2, \ldots, X_m)$.

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 $(U_1, U_2, \ldots, U_m)$ derived from the defined copula function (Equation (4)).

$$(X_1, X_2, \ldots, X_m) = (F_1^{-1}(U_1), F_2^{-1}(U_2), \ldots, F_m^{-1}(U_m)) \quad (4)$$

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.

$$\mathrm{C}(u_1, u_2, \ldots, u_m) = p[X_1 \le F_1^{-1}(u_1), X_2 \le F_2^{-1}(U_2), \ldots, X_m \le F_m^{-1}(U_m)] \quad (5)$$

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 ρ and Kendall's tau τ as given by Equation (6).

$$\boldsymbol{\rho}(\mathbf{x}, \mathbf{y}) = \sin(\tfrac{\pi}{2}\tau) \quad (6)$$

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

$$\mathrm{C}(u_1, u_2, \ldots, u_d) = \phi_R(\phi_1^{-1}(u_1), \phi_2^{-1}(u_2), \ldots, \phi_d^{-1}(u_d)) \quad (7)$$

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

$$\mathbf{C}(\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_d}) = \frac{1}{\sqrt{detR}} \exp(-\frac{1}{2}\begin{pmatrix}\phi_1^{-1}(u_1)\\ \vdots \\ \phi_d^{-1}(u_d)\end{pmatrix}^T * (\Sigma_d^{-1} - I) * \begin{pmatrix}\phi_1^{-1}(u_1)\\ \vdots \\ \phi_d^{-1}(u_d)\end{pmatrix} \quad (8)$$

Where, *I* denotes the $d \times d$ identity matrix.

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

$$\mathrm{C}(u_1, u_2, \ldots, u_d) = \int_{-\infty}^{t_\nu^{-1}(u_1)} \cdots \int_{-\infty}^{t_\nu^{-1}(u_1)} f_{t_1(\nu)(X)d_X} \quad (9)$$

Where, $f_{t_1(\nu)}$ denotes the *d*-dimensional Student-*t* density function with degree of freedom ν and t_ν^{-1} denotes the quantile function of a standard univariate Student-*t* distribution with degree of freedom ν.

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.

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*

Copula	Bivariate Copula ($C_\theta(a,b)$)	Parameter θ
Clayton	$[\max(a^{-\theta}+b^{-\theta}-1;0)]^{-1/\theta}$	$\theta \in [-1,\infty]\backslash 0$
Frank	$-\frac{1}{\theta}\log[1+\frac{(\exp(-\theta a)-1)(\exp(-\theta b)-1)}{\exp(-\theta)-1}]$	$\theta \in [-1,\infty]\backslash 0$
Gumbel	$\exp[-((-\log(a))^{\theta}+(-\log(b))^{\theta})^{1/\theta}]$	$\theta \in [1,\infty]$
Gaussian	$\frac{1}{\sqrt{detR}}\exp(-\frac{1}{2}\begin{pmatrix}\phi_1^{-1}(u_1)\\ \vdots \\ \phi_d^{-1}(u_d)\end{pmatrix}^T*(\Sigma_d^{-1}-I)*\begin{pmatrix}\phi_1^{-1}(u_1)\\ \vdots \\ \phi_d^{-1}(u_d)\end{pmatrix}$	$\theta \in [1,1]$
Student-t	$1-[(1-u)^{\theta}+(1-v)^{\theta}-(1-u)^{\theta}(1-v)^{\theta}]^{1/\theta}$	$\theta \in [1,\infty]$

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 θ 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.

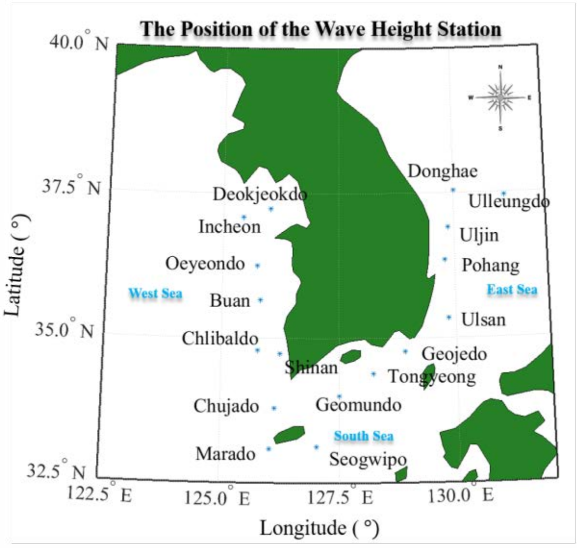

Figure 1. Location of the 17 wave height stations.

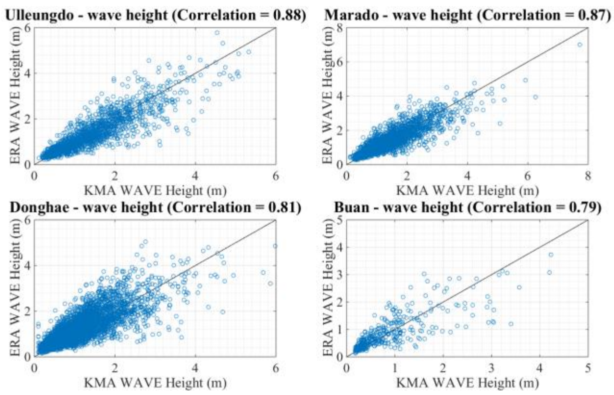

Figure 2. Scatter plots between ERA-Interim and KMA wave height.

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.

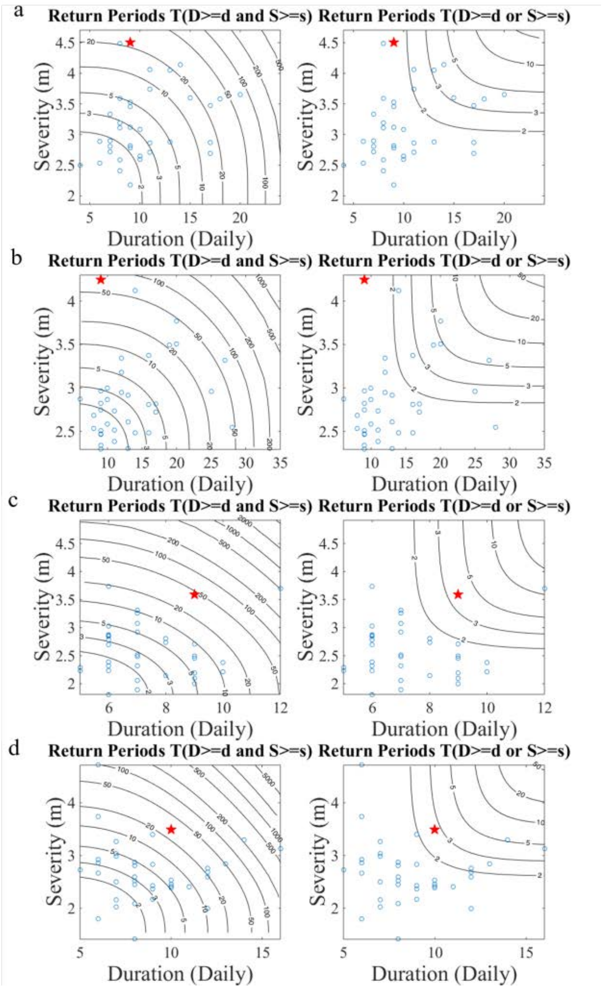

Figure 3: Bivariate return periods for wave height and period for four representative stations (stars indicate the recent extreme events) - (a: Ulleungdo, b: Donghae, c: Marado, d: Buan)

The bivariate frequency analysis was conducted for each station using the selected marginal probability density functions and Copula functions. The results showed that return periods for the extreme event that occurred in 2016 were higher than 20-years in Marado, Chujado and Seogwipo of Jeju Island, Ulleungdo and Donghae of the East Coast. On other hand, for the remaining 12 stations, the return periods for the extreme events that occurred in 2016 varied from 1.2 to 16.6 years. This indicates that most of the stations along the path of typhoons have a higher return period compared to the western coast of South Korea. Typhoons of South Korea mostly occur in the southwest and tend to move forward to the northwest (Kim *et al*., 2014).

Table 2. *Selected copula functions for each station*

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

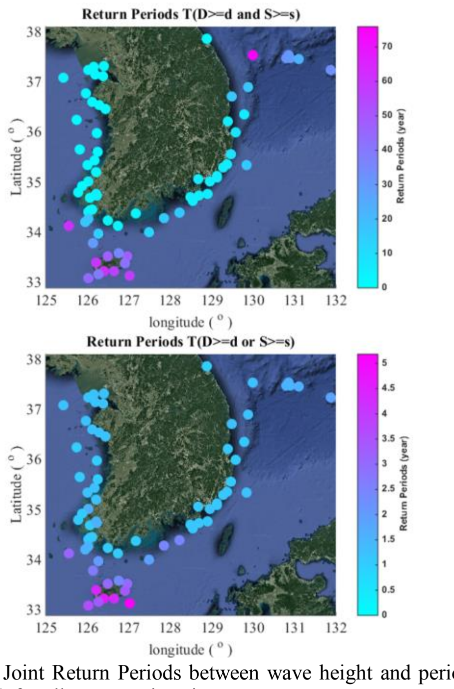

Figure 4. Joint Return Periods between wave height and period for the year-2016, for all wave stations in Korea.

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 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

Cho, K.W., and Maeng, J.H., 2007. Some Thoughts on Direction to Cope with the Sea level Rise in Korea, *The Korean Society for Marine Environmental Engineering*, Vol. 10, No. 4, pp. 227-234.

Cox, A.T., and Swail, V.R., 2001. A global wave hindcast over the period 1958–1997: validation and climate assessment. *Journal of Geophysical Research*: Oceans, Vol. 106, No. C2, pp. 2313-2329.

Durante.; Fabrizio., and Carlo Sempi., 2010. Copula theory: an introduction. Copula theory and its applications. *Springer* Berlin Heidelberg, pp. 3-31.

Hemer, M.A.; Wang, X.L.; Weisse, R., and Swail, V. R., 2012. Advancing wind-waves climate science: The COWCLIP project. *Bulletin of the American Meteorological Society*, Vol. 93, No. 6, pp. 791-796.

Kim, T.-J.; Kwon, H.-H., and Kim, K.-Y., 2014. Assessment of Typhoon Trajectories and Synoptic Pattern Based on Probabilistic Cluster Analysis for the Typhoons Affecting the Korean Peninsula. *J. KOREA WATER RESOURCES ASSOCIATION*, Vol. 47, No. 4, pp.385-396.

Korea Institute of Ocean Science & Technology., 2005. Deep sea design wave estimation report II. BSPE95100-1767-2, 450p.

Kwon, H.-H.; Lall, U., and Kim, S.J., 2016b. The unusual 2013–2015 drought in South Korea in the context of a multicentury precipitation record: Inferences from a nonstationary, multivariate, Bayesian copula model. *Geophysical Research Letters*, Vol. 43, pp.8534-8555, doi:10.1002/2016GL070270.

Kwon, H.-H., and Lall, U., 2016a. A copula-based nonstationary frequency analysis for the 2012–2015 drought in California. *Water Resources Research*, Vol. 52, pp.1-14, doi:10.1002/2016WR018959.

Lee, J.H.; Chung, G.H., and Kim, T.W., 2009. Evaluation of Flood Severity Using Bivariate Gumbel Mixed Model. *J. Korea Water Resources Association*, Vol. 42, No. 9, pp. 725-736.

Ling, C.H., 1965. Representation of associative functions. *Publicationes Mathematicae Debrecen*, Vol. 12, pp. 189-212.

Mastenbroek, C.; Burgers, G., and Janssen, P.A. E.M., 1993. The dynamical coupling of a wave model and a storm surge model through the atmospheric boundary layer. *Journal of Physical Oceanography*, Vol. 23, No. 8, pp. 1856-1866.

Prandle, D., and Wolf, J., 1978. The interaction of surge and tide in the North Sea and River Thames. *Geophysical Journal International*, Vol. 55, No. 1, pp. 203-216.

Sasaki, W.; Iwasaki, S. I.; Matsuura, T.; Iizuka, S., and Watabe, I., 2005. Changes in wave climate off Hiratsuka, Japan, as affected by storm activity over the western North Pacific. *Journal of Geophysical Research*: Oceans, Vol. 110, No. C9.

Schweizer, B., and Sklar, A., 1961. Associative function and statistical triangle inequalities. *Publicationes Mathematicae Debrecen*, Vol. 8, pp. 168-186.

Sklar, A., 1959. Fonctions de repartition an dimensions et leurs marges. *Université Paris*, Vol, 8, pp. 229-231.

Sterl, A.; Komen, G. J., and Cotton, P. D., 1998. Fifteen years of global wave hindcasts using winds from the European Centre for Medium-Range Weather Forecasts reanalysis: Validating the reanalyzed winds and assessing the wave climate. *Journal of Geophysical Research*: Oceans, Vol. 103, No. C3, pp. 5477-5492.

Wang, X.L., and Swail, V. R., 2001. Changes of extreme wave heights in Northern Hemisphere oceans and related atmospheric circulation regimes. *Journal of Climate*, Vol. 14, No. 10, pp. 2204-2221.

Wang, X.L., and Swail, V. R., 2002. Trends of Atlantic wave extremes as simulated in a 40-yr wave hindcast using kinematically reanalyzed wind fields. *Journal of climate*, Vol. 15, No. 9, pp. 1020-1035.

Yong, H.; Baoshu, Y.; Perrie, W., and Yijun, H., 2008. Responses of summertime extreme wave heights to local climate variations in the East China Sea. *Journal of Geophysical Research*: Oceans, Vol. 113, No. C9, pp. 1978-2012.

Yue, S., 2001. A bivariate gamma distribution for use in multivariate flood frequency analysis. *Hydrological Processes*, Vol. 15, No. 6, pp. 1033-1045.