JOINT RETURN VALUE ESTIMATION OF EXTREMES OF SIGNIFICANT WAVE HEIGHTS AND WIND SPEED IN COLOMBO, SRI LANKA

P. M. Senadeera[1], K. Perera[2]

[1]Postgraduate Institute of Science, University of Peradeniya, Sri Lanka, Email: manujayas@gmail.com
[2]Department of Engineering Mathematics, Faculty of Engineering, University of Peradeniya, Sri Lanka, Email: kanthip@pdn.ac.lk

ABSTRACT

Joint return period estimation of extreme wind speed and wave heights in Colombo area using Copulas is discussed in the study. Peaks over threshold method (POT) was used to find threshold values for the three hourly wind speed and wave heights. Mean Residual Life Plot (MRLP) and Generalized Pareto Distribution (GPD) were used to determine the threshold values for the data. All the data above threshold values were extracted as extreme values. Since there was a dependent structure between the two variables (the Kendall's rank correlation, p-value < 0.05), to obtain the joint distribution of the variables, Copula method was used. The copula combines the marginal distributions into a joint distribution. The advantage of the copula method is that no assumption is needed for the variables to be independent or normal or having same type of marginal distributions. Best marginal distribution and the best Copula were identified based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). To obtain the marginal distribution for the copula, Gamma, Normal, Lognormal, Weibull, Logistic and Exponential distributions were considered. The parameters of the distributions were estimated using the Maximum Likelihood method. Lognormal distribution was identified as the best marginal distribution based on AIC and BIC values for both extreme wind speed and wave heights. Five copula functions, Normal, Frank, Joe, Gumbel-Hougaard and the Clayton were used and identified that the Joe Copula was the best copula based on the AIC and BIC values. The parameters of the copulas were estimated using the itau method. Combining the identified marginal distributions by using the Joe Copula, the joint distribution of the extreme wind speed and sea wave heights was obtained. Return periods were calculated using the fitted joint distribution. The results show that for an extreme wind speed of 22.4 miles per hour with a corresponding wave height of 3.14 meters has a return period of 14 days.

Key words: Kendall's Tau, Joint distribution, Peaks over threshold, Copula, Return period

1. INTRODUCTION

With the end of more than 30 years of civil unrest, Sri Lanka is now on a path of rapid urbanization and development. As Colombo is the commercial capital and the largest city of Sri Lanka, many development projects are proposed to construct in it. Specially targeted area is Colombo coast. Because offshore and coastal structures are often exposed to extreme sea conditions and therefore their optimal design requires an estimation of return values of extreme wave heights and wind speed affecting them. Engineers will need these return periods to design the structures then they can use those details to build strong and long-lasting buildings.

Because extreme wind speed and wave heights do not follow the same probability distribution and there is a dependency between these variables the copula method was used. The advantage of the copula method is that no assumption is needed for the variables to be independent or normal or having the same type of marginal distributions[1]. The peaks-over-threshold (POT) method is a widely used approach for extreme value estimation that includes several of the largest order statistics exceeding a sufficiently high threshold in the available data [2]. Gumbel and Weibull probability distributions were fitted for extreme wave heights and wind speed and the best fitted empirical cumulative distribution was obtained using a least square method that minimizes the error between variables in the model [3].

2. METHODOLOGY

2.1. Data Used

Three hourly wind speed (MPH) and wave height (m) data for Colombo area from September, 2005 to September, 2007 obtained from the Department of Meteorology was used in this study.

2.2. Procedure for Analysis

In order to find the extreme values for the data, Peaks Over Threshold (POT) method was used to find the threshold values. These values were identified using the Mean Residual Life Plot and by considering the stability of the parameters of the Generalized Pareto Distribution. The data above the identified threshold values were extracted for the sample. Correlation between the extracted extreme values of wind speed anc wave heights was calculated using the Kendall's τ rank correlation. By fitting the Normal, Gamma, Weibull, Logistic, Lognormal and Exponential distributions and the copulas Normal, Clayton, Gumbel, Frank and Joe, the best marginal distributions for the two variables and the best copula were identified by using AIC and BIC values. By combining the identified marginal distributions using the best fitted copula, joint probability distribution was obtained. Using the fitted joint distribution, joint return periods were calculated.

2.3. Peaks Over Threshold (POT) Method

Conditional excess distribution can be approximate using distribution called, Generalized Pareto Distribution (GPD). This gives the probability of a random variable exceeding a given threshold value.

Distribution functions of GPD:

$$G(y; u, \acute{\sigma}, \xi) = 1 - \left[1 + \frac{\xi(y-u)}{\acute{\sigma}}\right]^{-\frac{1}{\xi}}$$

Where,

- $y - u > 0, 1 + \frac{\xi(y-u)}{\acute{\sigma}} > 0$
- Shape parameter of ξin GPD is same as Generalized Extreme Value(GEV) shape parameter
- Depends on parameter $\acute{\sigma}$ three parameter of GEV distribution as well as on the threshold value u, that is $\acute{\sigma} = \sigma + \xi(u - \mu)$
- When $\xi \to 0$ the GPD converges to Exponential family distribution

By considering the Mean Residual Life plot and the stability of the parameters of the Generalized Pareto Distribution, the threshold values would be selected.

2.4. Lognormal Distribution

A log-normal distribution is a continuous probability distribution of a random variable whose logarithm is normall distributed. Thus, if the random var is log-normally distributed, then $Y = log(x$ has a normal distribution. Like has a normal distribution, then $X = exp(Y)$ has a log-normal distribution. A random variable which is log-normally distributed takes only positive real values.

2.4.1 Probability Density Function

Probability density function is given by

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \text{ Where } x > 0$$

2.4.2. Cumulative Distribution Function

Cumulative distribution function is given by

$$F(x) = \frac{1}{2} + \frac{1}{2} erf\left[\frac{\ln x - \mu}{\sqrt{2}\sigma}\right]$$

2.5 Definition of Copula

Mathematically, a copula is a function which allows us to combine univariate distributions to joint distribution with a particular dependence structure [4], [5]. That is the copula is a multivariate distributionfunction with uniform marginal distributions. In the bivariate case, the copulas are the joint distribution of two random variables. The bivariate copula of two unit uniform random variables which defined as

$$C(u, v) = \Pr(U \leq u, V \leq v)$$

Let $H(x, y)$ be the joint distribution fu with arginal distribution functions $F_X(x)$ and $F_Y(y)$ which are umulative distribution functions of X and Y. Then there exists a function such that

$H(x, y) = C(F_X(x), F_Y(y))$ Where C is the copula.

To construct the copula,

Let $U = F_X(x)$ and $V = F_Y(y)$

$$C(u,v) = F\left(F_X^{-1}(u), F_Y^{-1}(v)\right)$$

This gives,

$$C\left(F_X(x), F_Y(y)\right) = F\left(F_X^{-1}(F_X(x)), F_Y^{-1}(F_Y(y))\right) = H(x,y)$$

The function $C(u,v)$ is called the copula.

2.5.1. Joe Copula

The Joe Copula function is given by;

$$C(u,v) = 1 - \left((1-u)^{\theta} + (1-v)^{\theta} - (1-u)^{\theta}(1-v)^{\theta}\right)^{\frac{1}{\theta}}$$

$$H(x,y) = C(F_X(x), F_Y(y))$$

Where $\theta > 0$ and with

$\varphi(t) = -log\left[1 - (1-t)^{\theta}\right]$ and

$\tau = 1 + \frac{4}{\theta^2}\int_0^1 t\,log(t)\left((1-t)^{\frac{2(1-\theta)}{\theta}}\right)dt$. The Joe copula covers only the positive dependency[6].

2.6. Joint Return Period

The joint return period is given by the formula

$$T_{X,Y}(x,y) = \left(\frac{1}{1 - F_{X,Y}(x,y)}\right)$$

Where, $F_{X,Y}(x,y)$ is the cumulative distribution function of the joint distribution function. The joint return period of an event e pres: :d by the equation represents that either X or Y or both values are exceeded.

3. RESULTS

A range for threshold values was selected from the linear range of mean residual life plot.

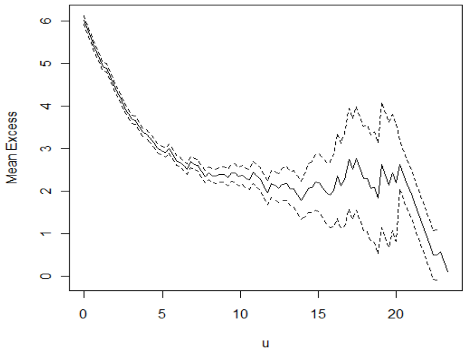

Figure 1: Mean Residual Life plot for Wind speed

For the wind speed data, the linear range is nearly between 5MPH and 10MPH. Further by fitting the GPD over the identified range and by considering the stability of the parameters, the threshold value was selected as 8MPH.

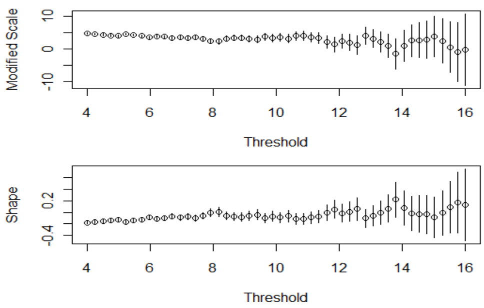

Figure 2: Fitted GPD for Wind Speed data

Similarly for the wave height data, the threshold value was selected as 1.5m by using the same procedure.

The Kendall's rank correlation (τ) was calculated for the extreme wind speed and wave heights. According to the Table 1, the rank correlation is significant (p-value < 0.05).

Table 1 : Kendall's rank correlation values

	Correlation	p- value
Kendall's rank correlation	(τ) 0.2190	< 0.05

According to Table 2, the minimum AIC values for extreme wind speed and wave heights are 288.44 and 52.08 respectively. Also the minimum BIC values for the extreme wind speed and wave heights are 292.53 and 56.17 respectively. These values were given by the lognormal distribution for both variables. Thus the Lognormal distribution is the best marginal distribution for extreme wind speed and also for the extreme wave heights.

Table 2 : AIC and BIC values for the fitted probability distributions

Probability Distributions	Wind Speed		Wave Height	
	AIC	BIC	AIC	BIC
Gamma	305.1	309.19	53.98	58.07
Exponential	393.29	395.33	196.54	198.59
Lognormal	288.44	292.53	52.08	56.17
Weibull	305.1	309.19	65.4	69.49
Normal	302.86	306.95	59.12	63.21
Logistic	301.98	306.06	60.9	64.99

Then, according to the Table 3, the smallest AIC and BIC values for fitted copula are 10.8029 and -8.8598 respectively. Both values were given by the Joe copula. Thus, the Joe copula is the best copula with the parameter $\theta = -1.4939$. By combining the identified marginal distributions and the Joe Copula, the joint probability distribution was obtained for the two variables.

Table 3 : Estimates of the parameters and the AIC, BIC values for the copulas

Copula	Parameter (θ)	AIC Value	BIC Value
Normal	0.3373	-3.372	-1.3290
Clayton	0.5610	2.798	4.8412
Gumbel	1.2800	-8.283	-6.239
Frank	2.0521	-3.856	-1.813
Joe	**-1.4939**	**-10.903**	**-8.860**

The copula based joint distribution is given by

$h(x, y)$

$$h(x,y) = 1 - \left(\left(\left(1 - F_X(x)\right)^{\theta} + \left(1 - F_Y(y)\right)^{\theta} - \left(1 - F_X(x)\right)^{\theta} \left(1 - F_Y(y)\right)^{\theta} \right) \right)^{\frac{1}{\theta}}$$

Where, $F_X(x)$ is the Cumulative Lognormal distribution of the wave heights with the mean 9 and the standard deviation 0.1852 and $F_Y(y)$ is the Cumulative Lognormal distribution of the extreme wind speed with the mean -2.3945 and the standard deviation 0.267.

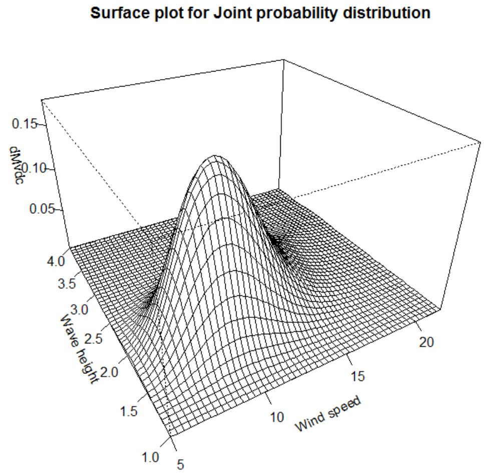

Figure 3: Surface plot for the joint probability distribution

Figure 3 displays the surface plot of the fitted joint probability density function, and the peak of the density plot at middle of the square indicates presence of strong positive dependence between extreme wind and wave heights. The return periods are longer when the extreme wind speed and wave heights are high. The pair of highest wind speed (22.4 MPH) and wave height (3.14 m) has the highest return period, which is about 14 days.

4. CONCLUSION

Copula based joint probability distribution is ideal for estimating joint return periods estimating for oceanographic variables such as extreme wind speed and wave heights because these variables are dependent and doesn't seem following a same distribution. Extreme values are found by determining relevant threshold values. It is used peaks over threshold method as the most suitable threshold value selection method in studying wind speed and wave height variables. Here for extreme wind speed and wave heights in Colombo area we propose Lognormal distribution as the best fitting marginal distribution. Also Joe copula is the best fitted copula for these variables. The return periods are longer when the extreme wind speed and wave heights are high. The pair of highest wind speed and the wave height has the highest return period, which is about 14 days. Moreover, theses return periods are useful when designing coastal structures.

5. REFERENCES

[1] L. Zhang, and V. P.Singh, "*Bivariate rainfall frequency distributions using Archimedean copulas",* Journal of Hydrology, 332(1), pp. 93-109, 2007.

[2] R. Nerzic, C. Frelin, and M. Prevosto, "*Joint distribution of wind/wave/current in West Africa and derivation of multivariate extreme I-Form contours*", Proceedings of the Sixteenth International Offshore and Polar Engineering Conference, Lisbon, Portugal, July 1-6, 2007, ISBN 978-1-880653-68-5, 2007.

[3] J. Pickands, "*Statistical inference using extreme order statistics*", Annals of Statistics 3, pp.119-131, 1975.

[4] G. G. Venter, "*Tails of copulas*", In Proceedings of the Casualty Actuarial Society (Vol. 89, No. 171, pp. 68-113), 2002.

[5] H. Manner, "*Estimation and model selection of copulas with an application to exchange rates",* METEOR, Maastricht research school of Economics of Technology and Organizations, 2007.

[6] R. B. Nelson, "*An Introduction to Copulas",* Springer-Verlag, 1999.