

# Dependency Modelling of Natural Rare Phenomena: Application on Oil Rigs

Hussein M. Badran<sup>1</sup>, Zaher Khraibani<sup>2,\*</sup>, Hussein Khraibani<sup>2</sup>,  
Hassan Zeineddine<sup>2</sup>, Aline Mefleh<sup>2</sup>, Hassan Hamie<sup>3</sup>

<sup>1</sup>Islamic University of Lebanon, Faculty of Economics and Business, Beirut, Lebanon

<sup>2</sup>Lebanese University, Faculty of Sciences, Department of Statistics, Hadath, Lebanon

<sup>3</sup>Lebanese Petroleum Administration, Beirut, Lebanon

**Abstract** Recently, the East Mediterranean region witnessed frequent extreme climate change conditions manifested in high waves, strong wind and seismic activity. This resulted in increased attention to the safety of existing oil rigs. Thus, based on the variable dependency of the extreme value theory, the study aims to propose a new method to control the effect of climate conditions, particularly the wind speed, which may inflict structural damage to oil rigs. Additionally, the study seeks to compare the extreme value dependence between two different regions in Lebanon in order to evaluate how oil rigs, which will be installed in Lebanon's maritime area, resist to high wind speed.

**Keywords** Extremal Dependence, Extreme Values, Oil Rigs, Wind speed, Mediteranea Sea

## 1. Introduction

Millions of barrels of oil equivalent are produced every day. In fact, at the end of 2015 production reached 97 million barrels per day [6]. Geological and geophysical studies conducted in the Lebanese offshore, including the interpretation of 2D and 3D seismic data, have revealed the presence of natural gas resources. Some prospects are located in water areas with a depth of 2,000 meters [11]. The fixed or floating platform shall be designed according to water depth and bathymetry of seabed to resist the extreme weather conditions. The future oil and gas exploration and production activities in the Lebanese offshore shall comply with the Health, Safety and Environment requirements in every phase of the value chain.

Nowadays, semi-submersible and dynamically positioned drilling ships are used in exploring and drilling exploratory wells in water depths down to 2,500 meters.

The cost to operate these rigs is enormous. Weather-related events can affect drilling operations by causing delays that can incur high costs. For this reason, the univariate extreme value theory has been highlighted in the research literature where the variables are independent and identically distributed [1], [7]. This theory is widely used widely in many disciplines such as hydrology, meteorology, biology, environment, flood, hurricane, finance and

insurance [2], [4], [5], [8], [23]. The latest studies have specifically focused on variable dependency [10], [12].

In this research paper, the interest is specifically oriented toward the study of the univariate extreme value theory as applied to wind speed, using the classical method of the Block Maxima and the Peak Over Threshold (POT) in order to calculate the return level of significant wind speed over periods of 50, 100, 500 and 1000 years to protect the design of floating semi-submersible. This return level is used by the engineers in the design of oil rigs, and for mooring system to avoid wind risks especially to floating rigs.

By using both the Frechet and Weibull distribution respectively in the block Maxima and POT methods, we demonstrate that the POT method is more adequate to study the return levels of the wind speed in Lebanon.

Furthermore, we are interested in studying the bivariate extreme value theory to compare the wind speed across two different regions in Lebanon: Jiyeh in Southern Lebanon and Tripoli in Northern Lebanon. The dataset used in this study is composed of daily wind speeds registered every six hours during the period of 9-11-2009 to 10-6-2015. The bivariate extreme value theory studies the dependence between two variables and calculates the extreme conditional quantiles through both the block maxima and POT methods. There are different models in the literature to estimate the parameters of the bivariate extreme distributions [13], [17]. In order to study the dependence of the wind speed in two different regions, we develop three models which are the logistic, asymmetric logistic in the block maxima method and Husler Reiss model in the POT method [15], [16].

\* Corresponding author:

Zaher.khraibani@gmail.com (Zaher Khraibani)

Published online at <http://journal.sapub.org/ijps>

Copyright © 2016 Scientific & Academic Publishing. All Rights Reserved

## 2. The Univariate Extreme Value Theory

The classical univariate extreme value theory was developed by Frechet, Fisher and Tippet [3]. Two methods are used for estimating the extreme quantiles [7], [9], [10]. The first one is the block maxima method, which suggests the use of the generalized extreme-value (GEV) to calculate the limit distribution of the maxima or minima based on the Fisher-Tippet theorem [14].

**Theorem:** Let  $X_1, \dots, X_n$  be a sequence of independent and identically distributed random variables, and  $M_n = \max(X_1, \dots, X_n)$ . If a sequence of pairs of real numbers  $(a_n, b_n)$  exists such that  $a_n > 0$  :

$$\lim_{x \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = G(x)$$

where  $G$  is a non degenerate distribution function, then the limit distribution  $G$  belongs to one of the three distributions below called the generalized extreme value *GEV* distributions whose standard forms are:

1-Gumbel:

$$\Omega(x) = \exp[-\exp(-x)]$$

Where  $x \in \mathbb{R}$ .

2-Frechet:

$$\phi_\theta(x) = \exp[-x^{-\theta}]$$

Where  $x \geq 0$  and  $\theta > 0$  is the shape parameter.

3-Weibull :

$$\psi_\theta(x) = \exp[-(-x)^\theta]$$

Where  $x \geq 0$  and  $\theta > 0$  is the shape parameter.

The GEV distribution is given by:

$$G_{\gamma, \Gamma, \sigma}(x) = G_\gamma\left(\frac{x - \Gamma}{\sigma}\right) = \exp\left[-\left(1 + \gamma\left(\frac{x - \Gamma}{\sigma}\right)\right)^{\frac{-1}{\gamma}}\right]$$

where  $x$  is such that  $1 + \gamma\left(\frac{x - \Gamma}{\sigma}\right) > 0$ ,  $\gamma \in \mathbb{R}$ ,  $\Gamma \in \mathbb{R}$ ,  $\sigma > 0$ .

$\Gamma$  the location parameter,  $\sigma$  the scale parameter and  $\gamma$  the shape parameter.

If  $\gamma \geq 0$ , we have a Frechet distribution.

If  $\gamma \leq 0$ , we have a Weibull distribution.

If  $\gamma = 0$ , we have a Gumbel distribution.

The second method is the *POT* method, which studies the values that exceed a given threshold and uses the Generalized Pareto distribution (*GPD*) whose survival function is the following:

$$\bar{H}_{\{\gamma, \sigma\}}(Y) = \begin{cases} \left(1 + \gamma \frac{y}{\sigma}\right)^{\frac{-1}{\gamma}} & \text{if } \gamma \neq 0 \\ \exp\left(\frac{-y}{\sigma}\right) & \text{if } \gamma = 0 \end{cases}$$

### Theorem [13]

There is an equivalence between the convergence in distribution of the maximum to a *GEV* and the convergence

in the distribution of the exceedances to a *GPD* [14]:

$$\lim_{x \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = G_\gamma(x) \\ \Rightarrow \lim_{u \rightarrow +x_F} \sup_{[0, x_F - u]} |\bar{F}_u(y) - \bar{H}_{\gamma, \sigma}(y)| = 0$$

Where  $x_F$  is the endpoint of the function  $F$ .

The advantage of the *POT* method is that it is easier to construct a sample of exceedances than a sample of maxima.

Using both methods we can calculate the return level over a period of  $T$  years of a given variable, which is the value of the variable that will be exceeded one time every  $T$  years with a probability of  $\frac{1}{T}$ .

### 2.1. Block Maxima Method and Parameters Estimation

The classical approach to modelling the extremal properties of an independent and identically distributed variables  $X_1, \dots, X_n$  is held by using the block maxima method, which relies on probabilistic results based on the Fisher and Tippet (1928) theorem [14]. To apply the block maxima method, we choose to consider a block of 15 days, and thus we have  $15 \times 4 = 60$  observations in each block; therefore, the year is divided into 24 blocks. Finally, we obtain 136 maxima that will constitute the studied sample.

The estimation of the GEV distribution parameter is obtained by using the maximum likelihood method [21] with a 95% confidence interval (CI) which is given in the following Table:

**Table 1.** Estimation of the GEV distribution parameters

|                    | $\Gamma$       | $\sigma$     | $\gamma$     |
|--------------------|----------------|--------------|--------------|
| Parameter          | 23.17          | 5.04         | 0.37         |
| Standard deviation | 0.52           | 0.46         | 0.10         |
| CI (95%)           | [22.14, 24.19] | [4.14, 5.95] | [0.17, 0.57] |

We find that  $\gamma > 0$ , so the distribution of the maximum is a Frechet distribution.

To verify this result we use the comparison models method. If we suppose the maxima to be independent and having the same GEV distribution, then for a very high value  $n$ , we have the deviance [19]:

$$D = 2l(\hat{\mu}, \hat{\sigma}) - l_p(\mu, \hat{\sigma}) \sim \chi_k^2$$

where  $l(\hat{\mu}, \hat{\sigma})$  is the maximum likelihood for the model with  $\gamma \neq 0$  and  $l_p(\mu, \hat{\sigma})$  the maximum likelihood for the model with  $\gamma = 0$ .

So, we consider the hypothesis  $H_0: \gamma = 0$ , we will reject  $H_0$  if  $D > C_\alpha$  where  $\alpha$  is the  $1 - \alpha$  quantile with  $\alpha = 5\%$ .

We obtain:

$$l(\hat{\mu}, \hat{\sigma}) = -462.888$$

and

$$l(0, \hat{\sigma}) = -471.323$$

Hence,

$$D = 2(-462.888 - (-471.323)) = 16.87$$

The value of  $D$  is higher to the critical value ( $C_\alpha = 3.84$ ), so we reject  $H_0$  and thus  $\gamma > 0$  and the GEV distribution is Frechet.

## 2.2. Return Level

The return levels for the periods of  $T$  years are given by the following formulas [19]:

$$Z_p = \bar{F}^{-1}(p) \simeq \begin{cases} \Gamma - \frac{\sigma}{\gamma} [1 - (-\log(1-p))^{-\gamma}] & \text{if } \gamma \neq 0 \\ \Gamma - \sigma \log(-\log(1-p)) & \text{if } \gamma = 0 \end{cases}$$

Next we search the value that will be exceeded once every  $n \times k \times T$  observations where  $p = \frac{1}{nkT}$  and  $n$  the number of observations per block,  $k$  the number of blocks per year.

If we consider the value that will be exceeded once every  $k \times T$  blocks then the return level of  $T$  years is:

$$x_q = \begin{cases} \hat{\Gamma} - \frac{\hat{\sigma}}{\hat{\gamma}} [1 - (-\log(1-q))^{-\hat{\gamma}}] & \text{if } \hat{\gamma} \neq 0 \\ \hat{\Gamma} - \hat{\sigma} \log(-\log(1-q)) & \text{if } \hat{\gamma} = 0 \end{cases}$$

The results obtained by the statistical software (R) are the following [18]:

**Table 2.** Estimations of the return levels  $Z_p$

| Return period (years) | 50                       | 100                       | 500                       | 1000                       |
|-----------------------|--------------------------|---------------------------|---------------------------|----------------------------|
| $q$                   | $\frac{1}{24 \times 50}$ | $\frac{1}{24 \times 100}$ | $\frac{1}{24 \times 500}$ | $\frac{1}{24 \times 1000}$ |
| $x_q$                 | 99.95                    | 102.19                    | 113.58                    | 117.60                     |

Based on Table 2, the wind speed that will be exceeded once every  $24 \times 50 = 1200, 2400, 12000, 24000$  blocks is 99.95, 102.19, 113.58, 117.60 km/h, respectively.

## 2.3. POT Method and Parameters Estimation

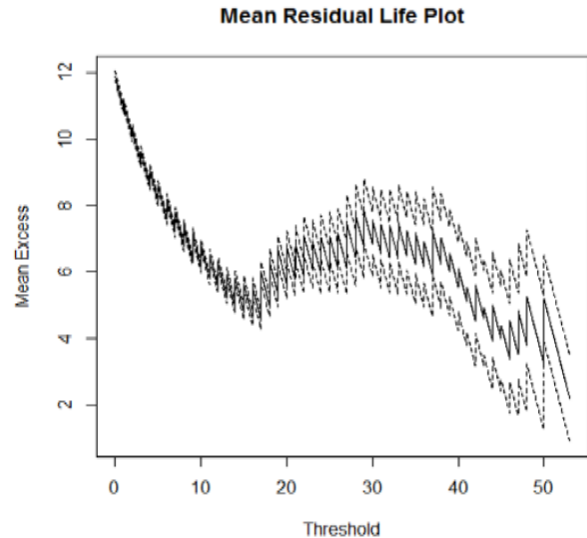
To fix a threshold  $u_0$  we consider all the values that exceed it by using the linear mean excess function (MEF), for  $u > u_0$  we have [22]:

$$MEF(u) = E(X - u | X > u) \text{ where } E(X) < \infty$$

With  $\gamma < 1$  we have:

$$E(X - u | X > u) = \frac{\sigma_u}{1 - \gamma} = \frac{\sigma_{u_0} + \gamma(u - u_0)}{1 - \gamma} = \frac{\gamma}{1 - \gamma} u + \frac{\sigma_{u_0} - \gamma u_0}{1 - \gamma}$$

Based on the mean residual life (*mrl*) plot in the software (R), we choose the value of the wind speed above by which the mean excess function becomes linear.



**Figure 1.** MRL plot

It seems that above 30 km/h the MEF becomes linear, so we choose 30 as a threshold and we have 145 exceedances among 8160 observations, which represent 1.78% of the observations.

The parameter's estimations are in the following Table:

**Table 3.** Parameters estimation by GPD

| Parameter          | $\sigma$      | $\gamma$       |
|--------------------|---------------|----------------|
| Estimation         | 9.63          | -0.27          |
| Standard deviation | 1.07          | 0.08           |
| CI (95%)           | [7.54, 11.73] | [-0.42, -0.12] |

We notice that 0 is not included in the confidence interval. To verify, we consider the below hypothesis:

$$H_0: \gamma = 0$$

We propose the Analysis of Variance (ANOVA) test:

**Table 4.** Parameter estimations by GPD

| Model | Test            | Deviance | T-test   |
|-------|-----------------|----------|----------|
| $M_1$ | $\gamma \neq 0$ | 868.2314 |          |
| $M_0$ | $\gamma = 0$    | 876.315  | 8.083595 |

The statistic test shows that  $8.083595 > 3.84$  (the 95% quantile of  $\chi^2_1$ ), so we reject  $H_0$  and hence the corresponding distribution is Weibull.

The return level for a period of  $T$  years is given by the following formula [22]:

$$Z_p = \bar{F}^{-1}(p) \simeq \begin{cases} u + \frac{\sigma}{\gamma} \left( \left[ \frac{p}{\bar{F}(u)} \right]^{-\gamma} - 1 \right) & \text{if } \gamma \neq 0 \\ u - \sigma \log \left( \frac{p}{\bar{F}(u)} \right) & \text{if } \gamma = 0 \end{cases}$$

where  $u$  is the threshold.

The obtained results are:

**Table 5.** Estimation of the return levels

| Return period (years) | 50                                 | 100                                 | 500                                 | 1000                                 |
|-----------------------|------------------------------------|-------------------------------------|-------------------------------------|--------------------------------------|
| $p$                   | $\frac{1}{50 \times 24 \times 60}$ | $\frac{1}{100 \times 24 \times 60}$ | $\frac{1}{500 \times 24 \times 60}$ | $\frac{1}{1000 \times 24 \times 60}$ |
| $Z_p$                 | 60.40                              | 61.27                               | 62.77                               | 63.23                                |

Based on Table 5 we can deduce that the wind speed be exceeded once every  $50 \times 24 \times 60 = 72000$ , 144000, 720000 and 1440000 observations is 60.40, 61.27, 62.77 and 63.23 km/h, respectively.

Finally, by comparing the results of the two methods in the case of univariate extreme values, we notice that there is a considerable difference due to the distributions in both methods being different. In fact, the Frechet distribution gives high and not very realistic return levels for the wind speed unlike the Weibull distribution, which gives more acceptable values.

Hence, we can say that the POT method is more appropriate than the block maximum one for the wind speed variable and, thus, the return levels of the wind speed in Jiyeh for periods of 50,100,500 and 1000 years are not very high and risky.

### 3. Bivariate Extreme Value Theory

The bivariate extreme value theory has so far received little attention in the literature [24]. As the univariate case, the bivariate extreme value is applied in several areas. For example, in the analysis of environmental extreme value data, there is a need of models of dependence between extremes from different sources. We study one more variable, in this section which is the wind speed in Tripoli to test the dependency between it and the wind speed in Jiyeh and then we calculate the conditional probabilities.

#### 3.1. Bivariate Block Maxima Method

The recorded database of the wind speed in Tripoli is similar to that of Jiyeh; we choose the same division of blocks. First we apply the logistic distribution method to estimate the parameters of the bivariate extreme distribution.

★ The bivariate logistic distribution function for a pair of variables  $(X, Y)$  where each one follows a univariate generalized extreme value distribution is [17]:

$$G(x, y) = \exp \left[ - \left( x^{\frac{-1}{\alpha}} + y^{\frac{-1}{\alpha}} \right)^{\alpha} \right]$$

where  $\alpha$  is the dependence parameter:

1. If  $\alpha \rightarrow 1$  then  $G(x, y) \rightarrow \exp \left[ - \left( \frac{1}{x} + \frac{1}{y} \right) \right]$  so we have an independence between the two variables.
2. If  $\alpha \rightarrow 0$  then  $G(x, y) \rightarrow \exp \left[ - \max \left( \frac{1}{x}, \frac{1}{y} \right) \right]$  so we have a perfect dependence.

The estimated parameters using the maximum likelihood

method are in the following Table:

**Table 6.** Estimation of parameters of case1 of the logistic model

| Parameter          | $\mu_1$       | $\sigma_1$  | $\gamma_1$  |
|--------------------|---------------|-------------|-------------|
| Estimation         | 23.28         | 5.11        | 0.31        |
| Standard deviation | 0.51          | 0.45        | 0.08        |
| CI (95%)           | [22.28,24.28] | [4.22,5.29] | [0.15,0.47] |

**Table 7.** Estimation of parameters of case 2 of the logistic model

| Parameter          | $\mu_1$       | $\sigma_1$  | $\gamma_1$  | $\alpha$    |
|--------------------|---------------|-------------|-------------|-------------|
| Estimation         | 31.68         | 8.76        | -0.07       | 0.38        |
| Standard deviation | 0.83          | 0.59        | 0.05        | 0.03        |
| CI (95%)           | [30.05,33.31] | [7.60,9.93] | [0.17,0.02] | [0.31,0.45] |

★ Secondly we apply the bivariate asymmetric logistic distribution function for the pair of variables  $(X, Y)$  [17]:

$$G(x, y) = \exp \left\{ - (1 - t_1)x^{-1} - (1 - t_2)y^{-1} - \left[ (t_1x^{-1})^{\frac{1}{\alpha}} + (t_2y^{-1})^{\frac{1}{\alpha}} \right]^{\alpha} \right\}$$

Where  $\alpha$  is the dependence parameter,  $t_1$  and  $t_2$  are the asymmetric parameters.

1. If  $t_1 = t_2 = 1$ , we have the logistic model.
2. If  $\alpha = 1$  or  $t_1 = 0$  or  $t_2 = 0$ , we have an independence.
3. If  $t_1 = t_2 = 1$  and  $\alpha \rightarrow 0$ , we have a perfect dependence.

The estimated parameters are:

**Table 8.** Estimation of parameters of case 1 of the asymmetric logistic model

| Parameter          | $\Gamma_1$   | $\sigma_1$  | $\gamma_1$  | $t_1$       |
|--------------------|--------------|-------------|-------------|-------------|
| Estimation         | 23.3         | 5.13        | 0.33        | 0.96        |
| Standard deviation | 0.51         | 0.46        | 0.09        | 0.09        |
| CI (95%)           | [22.29,24.3] | [4.23,6.03] | [0.16,0.49] | [0.78,1.14] |

**Table 9.** Estimation of parameters of case 2 of the asymmetric logistic model

| Parameter          | $\Gamma_2$    | $\sigma_2$    | $\gamma_2$      | $t_2$               | $\alpha$     |
|--------------------|---------------|---------------|-----------------|---------------------|--------------|
| Estimation         | 31.81         | 8.91          | -0.09           | 0.99                | 0.38         |
| Standard deviation | 0.86          | 0.63          | 0.05            | $21 \times 10^{-6}$ | 0.041        |
| CI(95%)            | [30.12, 33.5] | [7.67, 10.14] | [-0.19, 0.0005] | [0.9991, 0.9992]    | [0.29, 0.46] |

★ Third we apply the bivariate Husler-Reiss distribution function for a pair of variables  $(X, Y)$  [17]:

$$G(x, y) = \exp[-V(x, y)]$$

Where

$$V(x, y) = \frac{1}{x} \phi \left( \frac{a}{2} + \frac{1}{a} \log \frac{y}{x} \right) + \frac{1}{y} \phi \left( \frac{a}{2} + \frac{1}{a} \log \frac{x}{y} \right)$$

where  $x, y > 0$ ,  $\phi(\cdot)$  the normal cumulative distribution  $N(0,1)$ ;  $a$  represents the independence.

1. If  $a \rightarrow \infty$ ,  $V(x, y) \rightarrow \frac{1}{x} + \frac{1}{y}$  i.e we have an independence;
2. If  $a \rightarrow 0$ ,  $\frac{a}{2} + \frac{1}{a} \times \log \frac{y}{x} \rightarrow \text{sign} \left( \log \frac{y}{x} \right) \times \infty$  then  $V(x, y) \rightarrow \max \left( \frac{1}{x}, \frac{1}{y} \right)$  i.e total dependence.

The estimated parameters are:

**Table 10.** Estimation of parameters of case 1 of Husler-Reiss model

| Parameter            | $\Gamma_1$     | $\sigma_1$   | $\gamma_1$   |
|----------------------|----------------|--------------|--------------|
| Estimation           | 23.30          | 5.11         | 0.31         |
| Standard deviation e | 0.51           | 0.45         | 0.08         |
| CI (95%)             | [22.31, 24.30] | [4.23, 5.99] | [0.15, 0.46] |

**Table 11.** Estimation of parameters of case 2 of Husler-Reiss model

| Parameter          | $\Gamma_2$     | $\sigma_2$   | $\gamma_2$      | $a$          |
|--------------------|----------------|--------------|-----------------|--------------|
| Estimation         | 31.69          | 8.8          | -0.09           | 2.55         |
| Standard deviation | 0.83           | 0.60         | 0.05            | 0.26         |
| CI(95%)            | [30.06, 33.32] | [7.63, 9.97] | [-0.186, 9.973] | [2.04, 3.06] |

### 3.2. Models Comparison

To compare the three previous models and choose the most appropriate one we can use the Akaike information criterion (AIC). The best model is the one with the lowest AIC where  $AIC = 2k - 2L$  [17].

with  $k$ : the number of parameters and  $L$ : the log likelihood of the model.

**Table 12.** Models comparison

|                           | Standard deviation | AIC      |
|---------------------------|--------------------|----------|
| Logistic model            | 0.05               | 1779.658 |
| Asymmetric logistic model | ----               | 1784.596 |
| Husler-Reiss model        | 0.16               | 1778.723 |

**Table 13.** Verification of dependence test

|       | Deviance (D) | Degree of freedom | T-test | p-value               |
|-------|--------------|-------------------|--------|-----------------------|
| $M_1$ | 1765.65      | ---               | ---    | ---                   |
| $M_0$ | 1933.99      | 1                 | 168.34 | $1.7 \times 10^{-38}$ |

Since the logistic and Husler-Reiss models have the lowest and very similar AIC values, we choose the model with the lowest standard deviation of the dependence parameter, which is the logistic model.

To verify the dependence between the two variables, we propose the following statistic test:

$$H_0: \alpha = 1$$

$$H_1: \alpha \neq 1$$

We obtain a p-value  $< 5\%$ , so we strongly reject the independence hypothesis  $H_0$  and thus we conclude the dependency between the two regions in Lebanon.

### 3.3. Conditional Probability Quantile

We can also calculate the probability that the wind speed in Jiyeh exceeds a given value given that the wind speed in Tripoli exceeds 50 km/h:

$$P(X > z_p | Y > 50) = \frac{P(X > z_p, Y > 50)}{P(Y > 50)} \\ = \frac{1 - P(X < z_p) - P(Y < 75) + P(X < z_p, Y < 75)}{1 - P(Y < 75)}$$

We can simplify the calculation by transforming the marginal distribution of  $X$  and  $Y$  into a Frechet distribution. The transformation for  $Y$  is the following:

$$y_F = \left[ 1 + \gamma_y \left( \frac{50 - \mu_y}{\sigma_y} \right) \right]^{\frac{1}{\gamma_y}}$$

and for  $X$ :

$$z_p = x_F = \left[ 1 + \gamma_x \left( \frac{z_p - \mu_x}{\sigma_x} \right) \right]^{\frac{1}{\gamma_x}}$$

with the marginal probability of  $Y$ :

$$P(Y < y_F) = \exp \left( \frac{-1}{y_F} \right)$$

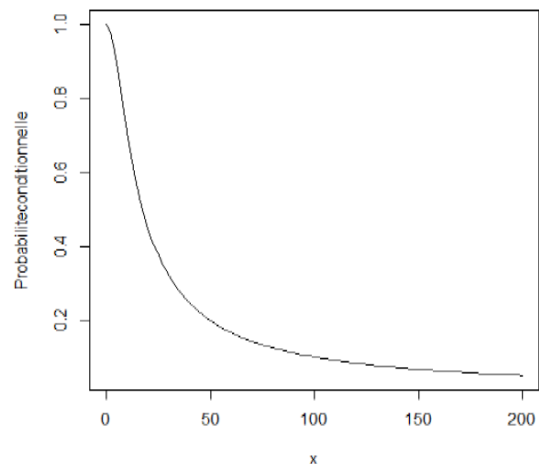
and of  $X$ :

$$P(X < z_p) = \exp \left( \frac{-1}{x_F} \right)$$

and

$$P(X < z_p, Y < 75) = G(x_F, y_F) = \exp \left[ - \left( x_F^{\frac{-1}{\alpha}} + y_F^{\frac{-1}{\alpha}} \right)^{\alpha} \right]$$

Thereby, we have the graph of the conditional probability:



**Figure 2.** Conditional probability

Finally, using this graph, we find that the probability that

the wind speed in Jiyeh exceeds 99.95 equal to  $\frac{1}{24 \times 50} = 8.33 \times 10^{-4}$  (return level of 50 years in the univariate case), given that the wind speed in Tripoli is higher than 50 km/h is 0.1.

### 3.4. Bivariate POT Method

As the univariate case, we choose the threshold for Tripoli wind speed data using the mrl plot [18].

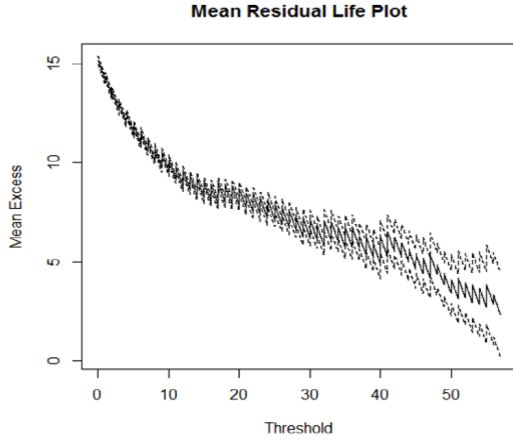


Figure 3. MRL Plot 2

By linearity, we choose  $u = 40$  above which there are 138 values that represent 1.69% of the total observations. The parameters are estimated using the censored likelihood method.

#### ★ Logistic GPD

Table 14. Estimation of parameters of case 1 of logistic model

| Parameter          | $\sigma_1$    | $\gamma_1$     |
|--------------------|---------------|----------------|
| Estimation         | 9.34          | -0.22          |
| Standard deviation | 0.98          | 0.07           |
| CI (95%)           | [7.42, 11.26] | [-0.36, -0.08] |

Table 15. Estimation of parameters of case 2 of the logistic model

| Parameter          | $\sigma_2$   | $\gamma_2$     | $\alpha$     |
|--------------------|--------------|----------------|--------------|
| Estimation         | 7.66         | -0.23          | 0.51         |
| Standard deviation | 0.98         | 0.07           | 0.03         |
| CI (95%)           | [6.23, 9.09] | [-0.34, -0.12] | [0.45, 0.57] |

#### ★ Asymmetric logistic General Pareto GPD

Table 16. Estimation of parameters of case 1 of the asymmetric logistic model

| Parameter          | $\sigma_1$    | $\gamma_1$     | $t_1$              |
|--------------------|---------------|----------------|--------------------|
| Estimation         | 9.65          | -0.23          | 1                  |
| Standard deviation | 1.17          | 0.09           | $2 \times 10^{-6}$ |
| CI (95%)           | [7.36, 11.95] | [-0.39, -0.06] | [0.99996, 0.99997] |

Table 17. Estimation of parameters of case 2 of the asymmetric logistic model

| Parameter          | $\sigma_2$   | $\gamma_2$     | $t_2$        | $\alpha$     |
|--------------------|--------------|----------------|--------------|--------------|
| Estimation         | 7.89         | -0.22          | 0.97         | 0.57         |
| Standard deviation | 0.88         | 0.077          | 0.10         | 0.04         |
| CI (95%)           | [6.17, 9.62] | [-0.37, -0.07] | [0.77, 1.17] | [0.49, 0.64] |

#### ★ Husler-Reiss GPD

Table 18. Estimation of parameters of case 1 of Husler Reiss model

| Parameter          | $\sigma_1$    | $\gamma_1$     |
|--------------------|---------------|----------------|
| Estimation         | 9.42          | -0.24          |
| Standard deviation | 0.95          | 0.06           |
| CI (95%)           | [7.56, 11.28] | [-0.36, -0.11] |

Table 19. Estimation of parameters of case 2 of Husler-Reiss model

| Parameter          | $\sigma_2$   | $\gamma_2$     | $a$          |
|--------------------|--------------|----------------|--------------|
| Estimation         | 7.56         | -0.22          | 1.77         |
| Standard deviation | 0.72         | 0.05           | 0.14         |
| CI (95%)           | [6.16, 8.96] | [-0.33, -0.12] | [1.49, 2.03] |

#### ★ Models comparison:

As before, the best model that will be chosen is the one with the lowest AIC [20].

Table 20. Models comparison

|                           | AIC             |
|---------------------------|-----------------|
| Logistic model            | 3875.009        |
| Asymmetric logistic model | 3887.713        |
| Husler-Reiss model        | <b>3869.492</b> |

We choose the Husler Reiss model since it has the lowest AIC.

To calculate the same conditional probability as before, we transform the marginal distribution of  $X$  and  $Y$  into a Frechet distribution.

The transformation for  $Y$  is:

$$y_F = - \left\{ \log \left[ 1 - \zeta_y \left[ 1 + \frac{\gamma_y (Y - u_y)}{\sigma_y} \right]^{\frac{-1}{\gamma_y}} \right] \right\}^{-1}$$

and for  $X$ :

$$z_p = x_F = - \left\{ \log \left[ 1 - \zeta_x \left[ 1 + \frac{\gamma_x (X - u_x)}{\sigma_x} \right]^{\frac{-1}{\gamma_x}} \right] \right\}^{-1}$$

where  $\zeta_x = \frac{r_x}{n}$  and  $r_x$  is the number of values that exceed the threshold of the  $X$  variable and  $n$  the total number of observations.

$\zeta_y = \frac{r_y}{n}$  where  $r_y$  is the number of values that exceed the threshold of  $Y$ .

So, the marginal probability of  $Y$  is:

$$P(Y < y_F) = \exp\left(\frac{-1}{y_F}\right)$$

of  $X$ :

$$P(X < z_p) = \exp\left(\frac{-1}{x_F}\right)$$

And,

$$\begin{aligned} P(X < z_p, Y < 50) &= G(x_F, y_F) \\ &= \exp\left\{-\left[\frac{1}{x_F}\phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{y_F}{x_F}\right) + \frac{1}{y_F}\phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{x_F}{y_F}\right)\right]\right\} \end{aligned}$$

where  $\phi(\cdot)$  the normal cumulative distribution  $N(0,1)$ .

Finally, the graph of the conditional probability  $P(X > z_p | Y > 50)$  is:

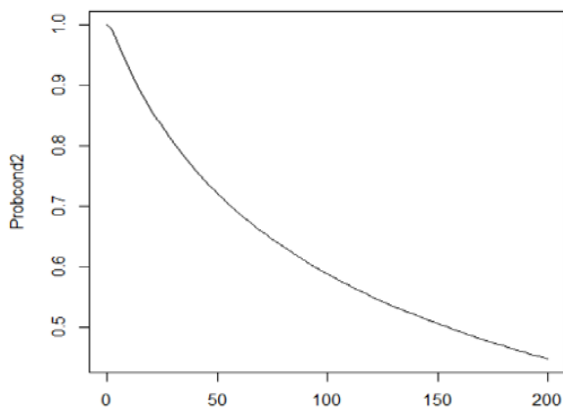


Figure 4. Conditional probability for case 2

Using this graph, we find that the probability for the wind speed at Jiyeh to exceed 60.4 km/h (return level of 50 years in the univariate case) given that the wind speed in Tripoli exceeds 50 km/h is 0.68.

This probability is high since the added condition is favorable and the difference between the two wind speeds is only 10.4 km/h (60.4-50).

## 4. Summary and Conclusions

The damage that can be caused by any extreme weather impact on an offshore rig in the future, will surely have huge burden on both the country and the operating companies working offshore Lebanon. The damage is not only limited to financial losses but also to the environmental and social impacts. Results have shown that the return levels of the wind speed in Jiyeh are not very high and risky.

The wind speed in Jiyeh and Tripoli dependent on each other and, thus, they almost represent the same risk level

since a high value in Tripoli will cause a high value in Jiyeh.

The return levels will then be used by engineers to forecast risks and protect the semi-submersible rigs from extreme wind speeds. The problem that remains is the choice of the return level to be used in the design of the rig, this choice will be a trade-off between the desired security level and the cost.

So finally, we believe that additional mathematical / statistical analysis should be conducted in future studies. To that end, modeling the return levels related to wave height, can help us evaluate another aspect of the structure integrity for potential offshore rigs.

## ACKNOWLEDGMENTS

We would like to acknowledge the Islamic University of Lebanon for the financial support of this research.

## REFERENCES

- [1] Coles S. (2001). An Introduction to Statistical Modeling of Extreme Values. Springer, London.
- [2] Abarbane, H.; Koonin, S.; Levine, H.; MacDonald, G.; Rothaus, O. (January 1992), "Statistics of Extreme Events with Application to Climate" (PDF), *JASON, JSR-90-30S*, retrieved 2015-03-03.
- [3] Haan, L. de and Ferreira, A. (2006). Extreme Value Theory: An Introduction. Springer Series in *Operations Research and Financial Engineering*, New York.
- [4] Embrechts P., Klüppelberg C. and Mikosch T. (1997). Modelling extremal events for insurance and finance. *Berlin: Springer Verlag*.
- [5] Novak S.Y. (2011). Extreme Value Methods with Applications to Finance. *Chapman & Hall/CRC Press, London. ISBN 978-1-4398-3574*.
- [6] International Energy Agency, I. (2015). *Oil Market Report*.
- [7] Lindgren, G.; Rootzen, H. (1987). "Extreme values: Theory and technical applications", *Scandinavian Journal of Statistics, Theory and Applications* 14: 241–279.
- [8] Buishand, T.A., de Haan, L. and Zhou, C. (2008). On spatial extremes: with application to a rainfall problem. *Annals of Applied Statistics* 2, 624–642.
- [9] Smith, R.L. and Weissman, I. (1994). Estimating the extremal index. *J.R. Statist. Soc. B* 56, 515–128.
- [10] Ledford, A.W. and Tawn, J.A. (1997). Statistics for near independence in multivariate extreme values. *Biometrika* 83, 169–187.
- [11] Hamieh, H. (2015). *Lebanese Petroleum Administration*.
- [12] Tawn, J.A. (1990). Modelling multivariate extreme value distributions. *Biometrika* 77, 245–253.
- [13] Pickands, J (1975). Statistical inference using extreme order statistics, *Annals of Statistics* 3: 119–131.

- [14] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004), *Statistics of Extremes: Theory and Applications*, *Wiley Series in Probability and Statistics*, Chichester, U.K.
- [15] Lindgren, G.; Rootzen, H. (1987). Extreme values: Theory and technical applications, *Scandinavian Journal of Statistics, Theory and Applications* 14: 241–279.
- [16] J. A. Tawn, *Bivariate Extreme Value Theory: Models and Estimation*. Oxford.
- [17] Ledford, A.W. and Tawn, J.A. (1996). Modelling dependence within joint tail regions, *J.R. Statist. Soc. B* 59, 475–499.
- [18] Coles, S.G.; Heffernan, J. & Tawn, J.A. (1999). Dependence measures for extreme value analyses, *Extremes*, 2, 339–365.
- [19] de Haan, L. & de Ronde, J. (1998). Sea and wind: multivariate extremes at work, *Extremes*, 1, 7–45.
- [20] Falk, M. & Michel, R. (2006). Testing for tail independence in extreme value models, *Ann. Inst. Statist. Math.*, 58, 261–290.
- [21] Ledford, A.W. and Tawn, J.A. (1997). Statistics for near independence in multivariate extreme values. *Biometrika* 83, 169 – 187.
- [22] Leadbetter, M. R. (1991). On a basis for 'Peaks over Threshold' modeling, *Statistics & Probability Letters* 12 (4): 357–362.
- [23] Smith, R.L. (2003). Statistics of extremes, with applications in environment, insurance and finance. Chapter 1 of *Extreme Values in Finance, Telecommunications and the Environment*, edited by B. Finkenstädt and H. Rootzén, *Chapman and Hall/CRC Press, London*, pp. 1–78.
- [24] Ramos, A. and Ledford, A. (2009). A new class of models for bivariate joint tails. *J.R. Statist. Soc. B* 71, 219–241.