Analysis between Temperature and Wind Speed in East Java using Bivariate Extreme Value Theory

A Oktaviarina and *A Sofro
Mathematics Department, Universitas Negeri Surabaya, Indonesia

*Corresponding author: ayuninsofro@unesa.ac.id

Abstract. Extreme weather now is the biggest effect of climate change. It is also a major issue over the past decade. High heat and hurricanes are some of its indicators. Both natural disasters have a big impact not only on humans but also on the presence of animals and the environment. Therefore, modelling of the disaster, especially for prediction extreme events is the basic need to minimize the damage and the losses. Both of the hazards come from extreme temperatures and wind speeds, respectively. Hence, this paper will analyze the issues using extreme value theory (EVT) approach. We will investigate it separately and then analyze both two cases simultaneously. The results show that using peaks over threshold univariate or separately approach is the best model among other approaches since providing the smallest AIC value. Meanwhile, the value of AIC for the block maxima is smaller than peaks over threshold using the bivariate approach.

1. Introduction
Drought and lack of clean water usually occurs in the long dry season due to extreme weather. In this season, the wind blows with a tendency to be high and air temperatures increase, see[1]. To anticipate the issue, further analysis needs to be done such as the prediction of maximum temperature and maximum wind speed. Using early warning information, we believe that the community can prepare and minimize the effect of the hazards.

One of the appropriate approaches for extreme phenomena is the extremes value theory (EVT). Extreme value theory is a branch of a statistical method for conducting extraordinary cases. Block maxima and peak over threshold are some methods in EVT, see [2]. The researchers have been interested in the research such as application in passing maneuvers, see [3]. Tansarthe has predicted the return level of hot days in the Persian Gulf using the Peak Over Threshold method, see [4].

In this case, between temperature and speed wind has correlation. Therefore, the most appropriate method is using bivariate extremes value theory. Some studies about these have been developed. Chen Wang has developed to obtain joint crash probability, see [5]. Application in different traffic conflict indicators for road safety estimation and actual crash data can also be found in [6].

In this paper, we will focus on modeling temperature and wind speed. We will provide the models and the return level.
2. Method

2.1 Extreme Value Theory
Extreme value theory concerns with phenomena of extreme data. The method has two approaches which are block maximums and peaks over threshold. EVT uses a single process is called univariate EVT, see [7].

2.2 Univariate Generalized Extreme Value (UGEV) Distribution
Let $X_1, X_2, \ldots, X_n$ is a independent random variables sequence with a common distribution function $F(x)$, and let $M_n = \max\{X_1, X_2, \ldots, X_n\}$ with $P(M_n \leq y) = F^n(y)$ correspondents to Block Maximum. When $n \to \infty, M_n$ will converge to a univariate generalized extreme value (UGEV) distribution:

$$F(\mu, \sigma, \xi) = \exp\left[-\left(1 + \xi \left(\frac{y-\mu}{\sigma}\right)\right)^{-1/\xi}\right], \xi \neq 0$$

(2.1)

and probability density function for UGEV as follows

$$f(\mu, \sigma, \xi) = \frac{1}{\sigma}\left[1 + \xi \left(\frac{y-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}-1}\exp\left[-\left(1 + \xi \left(\frac{y-\mu}{\sigma}\right)\right)^{-1/\xi}\right], \xi \neq 0$$

(2.2)

where $(\mu, \sigma, \xi)$ are the location, scale and shape parameters respectively. The Frechet case is obtained when $\xi > 0$, the negative Weibull when $\xi < 0$. The Gumbel case is defined by continuity when $\xi \to 0$, see [8].

The univariate GEV parameters can be estimated using maximum likelihood estimation (MLE), see [10]. The process of MLE for univariate GEV using equation (2.2) as follows

$$L(\mu, \sigma, \xi|x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i; \mu, \sigma, \xi)$$

$$L(\mu, \sigma, \xi) = \prod_{i=1}^{n} \frac{1}{\sigma}\left[1 + \xi \left(\frac{y_i-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}-1}\exp\left[-\left(1 + \xi \left(\frac{y_i-\mu}{\sigma}\right)\right)^{-1/\xi}\right]$$

$$\ln[L(\mu, \sigma, \xi)] = \ln \left[\prod_{i=1}^{n} \frac{1}{\sigma}\left[1 + \xi \left(\frac{y_i-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}-1}\exp\left[-\left(1 + \xi \left(\frac{y_i-\mu}{\sigma}\right)\right)^{-1/\xi}\right]\right].$$

2.3 Univariate Generalized Pareto (UGP) Distribution
The univariate generalized Pareto distribution for fitting peak over threshold method which has the function

$$F(z; \sigma_u, \xi) = 1 - \left(1 + \xi \left(\frac{z-\mu}{\sigma_u}\right)\right)^{-1/\xi}, \xi \neq 0$$

(2.3)

and probability distribution function for UGP is

$$f(z; \sigma_u, \xi) = \frac{1}{\sigma_u}\left[1 + \xi \left(\frac{z-\mu}{\sigma_u}\right)\right]^{-\frac{1}{\xi}-1}, \xi \neq 0$$

(2.4)

where $u$ is a high threshold, see [9].
The univariate GEV parameters can be estimated using MLE, see [10]. The procedure of MLE for univariate GPD using equation (2.4) is

\[
L(\sigma, \xi | x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i; \mu, \sigma, \xi)
\]

\[
L(\sigma, \xi) = \prod_{i=1}^{n} \frac{1}{\sigma_u} \left(1 + \xi \left(\frac{z_i - u}{\sigma_u}\right)\right)^{-\frac{1}{\xi}}
\]

\[
\ln[L(\mu, \sigma, \xi)] = \ln \left[ \prod_{i=1}^{n} \frac{1}{\sigma_u} \left(1 + \xi \left(\frac{z_i - u}{\sigma_u}\right)\right)^{-\frac{1}{\xi}} \right]
\]

2.4 Bivariate Generalized Extreme Value (BGEV)

Bivariate Generalized Extreme Value distribution function is

\[
G(x, y) = \exp\{-V(\bar{x}, \bar{y})\}, \quad \bar{x} > 0, \quad \bar{y} > 0
\]

where \(V(\bar{x}, \bar{y}) = 2 \int_0^1 \max \left(\frac{x}{\bar{x}}, \frac{1-w}{\bar{y}}\right) dH(w)\) and \(H\) is a distribution function on \(w \in [0,1]\) satisfying the mean constraint \(\int_0^1 w dH(w) = 1/2\) [7].

The BGEV parameters can be estimated using MLE. The procedure can be found in [7].

2.5 Bivariate Generalized Pareto (BGP)

Bivariate Generalized Extreme Value distribution function is

\[
G(x, y) = \exp\{-V(\bar{x}, \bar{y})\}, \quad x > u_x, \quad y > u_y
\]

where \(V(\bar{x}, \bar{y}) = 2 \int_0^1 \max \left(\frac{x}{\bar{x}}, \frac{1-w}{\bar{y}}\right) dH(w)\) and \(H\) is a distribution function on \(w \in [0,1]\) satisfying the mean constraint \(\int_0^1 w dH(w) = 1/2\), see [7].

Censored likelihood estimation function for Bivariate GPD is obtained as

\[
L(\theta) = \prod_{i=1}^{n} \psi(\theta, (x_i, y_i))
\]

where \(\theta\) denotes the parameters of \(G\) and

\[
\psi(\theta, (x_i, y_i)) = \begin{cases} 
\frac{\partial^2 F}{\partial x \partial y}|_{(x,y)}, & \text{if } (x, y) \in R_{1,1} \\
\frac{\partial G}{\partial x}|_{(x,u_y)}, & \text{if } (x, y) \in R_{1,0} \\
\frac{\partial G}{\partial y}|_{(u_x,y)}, & \text{if } (x, y) \in R_{0,1} \\
\end{cases}
\]

otherwise \(\psi(\theta, (x_i, y_i)) = G(u_x, u_y)\), if \((x, y) \in R_{0,0}\), see [12].

3. Results and Discussion

In this paper, we start to analyze using univariate extreme value theory using block maxima and peak over threshold. Then, we analyze using bivariate extreme value theory using block maxima and peak over threshold. The last step is to determine the best model for data based on AIC value.
The data consist of temperature maximum and wind speed maximum. The total observation is 365 daily data. The lowest wind speed data is 2 m/s and the highest is 9 m/s. The lowest temperature data is 24°C and the highest one is 37.2°C.

As we mentioned previous, the data is modelled using the block maxima univariate method. Estimated parameters of the block maxima method are given as follow

| No | Coefficients | Temperature | Wind Speed |
|----|--------------|-------------|------------|
|    | Estimate     | Std Error   | Estimate   | Std Error   |
| 1  | Location ($\mu$) | 33.29       | 4.0        | 0.08        |
| 2  | Scale ($\sigma$) | 1.45        | 1.33       | 0.05        |
| 3  | Shape ($\xi$) | -0.35       | -0.17      | 0.03        |

Table 1 shows that data follow generalize extreme value (GEV) distribution, especially Weibull distribution because all the shape parameter have negative value, see [13].

The univariate GEV distribution result from Table 1 is

$$f(y; \mu = 33.29, \sigma = 1.45, \xi = -0.35) = \frac{1}{1.45} \left[ 1 - 0.35 \left( \frac{y - 33.29}{1.45} \right) \right]^{1.35 - 1}, \xi \neq 0$$

Then, the data is modelled with peaks over threshold method with threshold=33 for temperature and threshold=3 for wind speed. In Table 2, it shows that data has generalized Pareto distribution. The detail of estimated parameters is shown in Table 2.

| No | Coefficients | Temperature | Wind Speed |
|----|--------------|-------------|------------|
|    | Estimate     | Std Error   | Estimate   | Std Error   |
| 1  | Location ($\mu$) | -           | -          | -          |
| 2  | Scale ($\sigma$) | 1.90        | 3.04       | 0.19       |
| 3  | Shape ($\xi$) | -0.43       | -0.49      | 0.036      |

The peaks over threshold univariate model can be served as follows

$$f(z; \sigma_u = 1.9, \xi = -0.43) = \frac{1}{1.9} \left( 1 - 0.43 \left( \frac{z - u}{1.9} \right) \right)^{-1 + \frac{1}{0.43}}, \xi \neq 0$$

Data is modeled using the bivariate block maxima methods. Table 3 that the data have generalized extreme value and e Weibull distribution because the shape parameter had a negative sign.
Table 3. Temperature and Wind speed Parameter Estimation of Block Maxima Bivariate

| No | Coefficient   | Estimate | Std Errors | Coefficient   | Estimate | Std Errors |
|----|---------------|----------|------------|---------------|----------|------------|
| 1  | Location (μ)  | 4.02     | 0.08       | Location (μ)  | 33.29    | 0.08       |
| 2  | Scale (σ) 1  | 1.34     | 0.05       | Scale (σ) 2  | 1.46     | 0.05       |
| 3  | Shape (ξ) 1  | -0.15    | 0.04       | Shape (ξ) 2  | -0.35    | 0.02       |

The block maxima bivariate model follows bivariate generalized extreme value (BGEV) distribution, so the distribution is

\[ G(x, y) = \exp[-V(\tilde{x}, \tilde{y})], \quad x > u_x, y > u_y \]

where

\[ V(\tilde{x}, \tilde{y}) = \left( 1 - 1.59 \left( \frac{x}{5.57} \right)^{1/1.59} \right) \left( 1 - 0.79 \left( \frac{y}{2.43} \right)^{1/0.79} \right) \]

Using threshold=3.33 for modeling data with peaks over threshold method. The result is given in Table 4 below:

Table 4. Temperature and Wind speed Parameter Estimation of Peaks Over Threshold Bivariate

| No | Parameter | Estimate | Parameter | Estimate |
|----|-----------|----------|-----------|----------|
| 1  | Location (μ) 1 | -        | Location (μ) 2 | -        |
| 2  | Scale (σ) 1  | 5.57     | Scale (σ) 2  | 2.43     |
| 3  | Shape (ξ) 1  | -1.59    | Shape (ξ) 2  | -0.79    |

The model of bivariate generalized pareto distribution based on in Table 4 is

\[ G(x, y) = \exp[-V(\tilde{x}, \tilde{y})], \quad x > u_x, y > u_y \]

where

\[ V(\tilde{x}, \tilde{y}) = \left( 1 - 1.59 \left( \frac{x-3.33}{5.57} \right)^{-1+1/1.59} \right) \left( 1 - 0.79 \left( \frac{y-3.33}{2.43} \right)^{-1+1/0.79} \right) \]

The last step is determining the best model. From Table 5, we can see that block maxima bivariate model is better than peaks over threshold which has AIC=2541.71.

Table 5. AIC values of different Models

| No | Model                  | AIC      |
|----|------------------------|----------|
| 1  | Univariate Block Maxima| 2552.28  |
| 2  | Univariate POT         | 1521.65  |
| 3  | Bivariate Block Maxima | 2541.71  |
| 4  | Bivariate POT          | 2000010  |

4. Conclusion

Temperature and wind speed data are modelled by univariate and bivariate model. The peaks over threshold univariate is the best model due to providing the smallest AIC value.
5. Acknowledgements
We would like to thank Universitas Negeri Surabaya, Indonesia, which give us a grant and opportunity to disseminate an intellectual of our research.

References

[1]. Gamage A U, Pearson D and Hanna F 2016 South East Asia J. of Public Health 6 3
[2]. Coles S G 2001 An introduction to Statistical Modeling of Extreme Values (Switzerland: Springer)
[3]. Farah H and Azevedo L 2017 IATSS Res. 41 12
[4]. Tanarhte M, Hadjinicolaou P, and Lelieveld J 2015 Clim. Res. 63 99
[5]. Wang C, Xu C, and Dai Y 2019 Accid. Anal. Prev. 123 365
[6]. Zheng L, Sayed T, and Essa M 2018 Acc. Anal. & Prev. 123 314
[7]. Zheng L and Sayed T 2019 Transp. Res. 103 211
[8]. Ribatet M 2011 A User’ s Guide to the POT Package (Version 1. 4) An Introduction to the EVT the univariate case
[9]. Gilleland E and Katz R W 2016 J. Stat. Softw. 72
[10]. Oktaviarina A, Sofro A, and Sutanto H 2018 Proc. of the Int. Conf. on Sci. and Tech. 182018.170
[11]. Sciences H, Journal J, June H, Sandoval C E and Raynal-villaseñor J 2008 Hydro. Sci. J. 53
[12]. Zheng L, Ismail K, Sayed T, and Fatema T, 2018 Accid. Anal. Prev. 120 83
[13]. Kim Y, Jang S C, and Lim T J 2015 Nucl. Eng. Technol. 1 59