Identification of Climate Change with Generalized Extreme Value (GEV) Distribution Approach

Anita Rahayu
Department of Statistics, Bina Nusantara University
Jalan KH. Syahdan No. 9 Palmerah, Jakarta 11480, Indonesia
E-mail: n33_tha@yahoo.co.id

Abstract. Some events are difficult to avoid and gives considerable influence to humans and the environment is extreme weather and climate change. Many of the problems that require knowledge about the behavior of extreme values and one of the methods used are the Extreme Value Theory (EVT). EVT used to draw up reliable systems in a variety of conditions, so as to minimize the risk of a major disaster. There are two methods for identifying extreme value, Block Maxima with Generalized Extreme Value (GEV) distribution approach and Peaks over Threshold (POT) with Generalized Pareto Distribution (GPD) approach. This research in Indramayu with January 1961-December 2003 period, the method used is Block Maxima with GEV distribution approach. The result showed that there is no climate change in Indramayu with January 1961-December 2003 period.

1. Introduction
As the vast archipelago country, Indonesia was strongly influenced by global climate change, therefore, Indonesia was able to predict the impacts of climate variability and change and rising sea levels. That ability is needed to perform the development in various aspects of life such as agriculture, fisheries, forestry, urban with infrastructure, health, coastal wetlands, and small islands. That ability is also required so Indonesia can prepare mitigation measures, anticipation, and adaptation to minimize the negative impacts of climate variability and change. One method that can be used to minimize that impact is Extreme Value Theory (EVT). There are two methods in identifying extreme value, Block Maxima with Generalized Extreme Value (GEV) distribution approach and Peaks over Threshold (POT) with Generalized Pareto Distribution (GPD) approach. Some studies have used the EVT method that is applied in various fields, including: economy [1, 2, 3, 4, 5, 6], telecommunications [7, 8, 9, 10], as well as the hydrology and climatology [11, 12, 13, 14, 15, 16, 17, 18, 19, 20].

As an island nation located in tropics, Indonesia is one of the country’s most vulnerable to the threat and impact of extreme events. The purpose of doing this study is to apply EVT in identifying climate change in Indramayu.

2. Extreme Value Theory (EVT)
Extreme events that often occur in the areas of insurance, economics, climatology, hydrology, and telecommunications are indicated by the presence of an observation that is very high (maximum) and very low (minimum). The interesting thing is to determine the probability (maximum and minimum)
of rare occurrence (tail distribution). One of the statistical methods used to study the behavior of the tail distribution is the Extreme Value Theory (EVT). There are two methods in identifying movements of the extreme value are taking the maximum value in a given period (called Block Maxima method) and retrieve the values that pass through a given threshold (called Peaks Over Threshold method) [21].

2.1 Block Maxima Method

One of the methods in identifying movements of the extreme value is Block Maxima, which identifies the extreme value based on the value of maximum observation data are grouped according to certain periods. In this method, observation data is divided into blocks in a certain period, example monthly, quarterly, semester, or year. Then for each block is determined the magnitude of observation data and the value is the maximum value of the extremes for each block and were used as the sample. Figure 1 shows that the observations which were used as the sample is $x_1, x_2, x_3, x_4, x_5$.

![Figure 1. Illustration of Taking Sample Data Precipitation with Block Maxima](image)

Generalized Extreme Value (GEV) distribution is a family of continuous distribution is built into the EVT to combine the Gumbel, Frechet, and Weibull distribution, known as the extreme value distribution of type I, II, and III. On the GEV distribution, $X$ is the random variable which has Probability Density Function (PDF) is as follows.

$$f(x; \mu, \sigma, \xi) = \begin{cases} 
\frac{1}{\sigma} \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} - \left[ 1 + \xi \left( \frac{X - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}}, & \xi \neq 0 \\
\frac{1}{\sigma} \exp \left\{ -\left( \frac{x - \mu}{\sigma} \right) \right\} \exp \left\{ -\exp \left( \frac{x - \mu}{\sigma} \right) \right\}, & \xi = 0
\end{cases}$$

(1)

Where a parameter of location is $\mu$, a parameter of scale is $\sigma$, and a parameter of shape is $\xi$. Shape parameter $\xi$ determines the behavior of the tail distribution. Distribution type defined with $\xi = 0$, $\xi > 0$, and $\xi < 0$ and can be likened to the Gumbel, Frechet, and Weibull distribution.
3. Parameter Estimation of Generalized Extreme Value (GEV) Distribution

Parameter estimation of Generalized Extreme Value (GEV) distribution can be done with Maximum Likelihood Estimation (MLE) using the following stages.

a. Take a random sample of \( n \) \( x_1, x_2, \ldots, x_n \)

b. Make likelihood function

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

(2)

c. Likelihood function utility by making ln from equation (2)

Qualified need \( \frac{\partial \ln L(\theta)}{\partial \theta} = 0 \) so obtained \( \hat{\theta} \)

Qualified enough \( \frac{\partial^2 \ln L(\theta)}{\partial \theta \partial \theta^T} = H(\theta) \) called hessian matrix

\( \hat{\theta} \) Maximize \( L(\theta) \) the definite positive terms \( H(\hat{\theta}) \)

If the results obtained from the first derivative of the ln function likelihood against each of these parameters is not closed form then required further numerical analysis to complete the equation.

4. Return Level

The return level is the maximum value which expected to be reached within the period of time one \( k \) with period \( p \), or in other words, in the \( k \) and \( p \) period, precipitation will reach the maximum value \( R^k_p \) one time [2]. The formula to estimate return levels are:

\[
\hat{R}^k_p = \begin{cases} 
\hat{\mu} - \frac{\sigma}{\xi} \left\{ 1 - \ln \left( \frac{1}{k} \right) \right\}^{-\frac{1}{\xi}}, & \xi \neq 0 \\
\hat{\mu} - \hat{\sigma} \ln \left\{ -\ln \left( \frac{1}{k} \right) \right\}, & \xi = 0 
\end{cases}
\]

(3)

5. Method

The data used in this research is dasarian rainfall data in Indramayu station during the period January 1961-December 2003. The data obtained from the Agency of Meteorology Climatology and Geophysics (BMKG). The steps of the research methods used are:

a. Create description of dasarian rainfall data in Indramayustation with 1961-2003 periods

b. Identifying the existence of a fat tail distribution
c. Separates the data into two periods, namely, the period I (1961-1990) and period II (1991-2003)
d. Identifying extreme values using Block Maxima method, which compiled the rainfall data based on block 3 monthly for each period, that is, December-January-February (DJF), March-April-May (MAM), June-July-August (JJA), and September-October-November (SON)
e. Identify the distribution pattern of dasarian rainfall data in 1961-2003
f. Identify the distribution pattern of rainfall data for each period

g. Perform parameter estimation for each period with MLE method, as well as create a confidence interval for each estimate parameters that have been obtained from the MLE method
h. Do hypothesis testing for each parameter estimation obtained from the MLE method using normal approach
Evaluating the suitability of the data distribution pattern of theoretical distribution pattern
Calculating the estimate return levels

6. Apply Extreme Value Theory (EVT) in Identifying Climate Change in Indramayu
Rainfall pattern in Indramayu station with 1961-2003 method is monsoon which shown with a plot of average monthly rainfall pattern resembles the letter U.

![Figure 2. Rainfall Histogram in Indramayu Station with 1961-2003 Period](image)

Figure 2 shows that the distribution pattern of dasarian rainfall data in Indramayu station with 1961-2003 period had tail distribution that goes down slowly when compared to a normal distribution, as a result of the extreme value of the opportunities will be greater than with a normal distribution modeling or commonly referred to as a fat-tailed distribution (heavy tail). With the indications on this research can use the EVT with extreme data retrieval using Block Maxima method.

| Parameter | Period I (1961-1990) | Period II (1991-2003) |
|-----------|----------------------|-----------------------|
| \( \hat{\mu} \)    | 96 [85; 108]          | 96 [77; 115]          |
| \( \hat{\sigma} \)  | 63[54; 71]            | 66[52; 81]            |
| \( \hat{\xi} \)     | 0.1 [-0.01; 0.6]      | 0.1 [-0.2; 0.8]       |

On Block Maxima method, dasarian rainfall data in Indramayu station with 1961-2003 periods are divided into 3 monthly blocks for each period. The block is the DJF, MAM, JJA, and SON. Then for each block is determined the magnitude of the maximum rainfall data and that data is extremes data for each block and were used as the sample. With the Block Maxima method obtained 172 extreme data in 1961-2003, with details of the 120 extreme data for the period I (1961-1990) and 52 extreme data for the period II (1991-2003).

The location parameter \( \mu \) indicates the location of the point from data centralization, the scale parameter \( \sigma \) indicates the pattern of diversity data, and shape parameter \( \xi \) indicates the right end point behavior of functions there. Based on Table 1 visible that the value of the location parameter
estimation $\mu$ obtained by the MLE method for period I and II with the centralization of data is 96 mm. Additionally the value of each parameter estimation obtained by the MLE method is lies in confidence interval of 95%.

After obtained a value for the parameter estimation $\mu, \sigma,$ and $\xi$, the next step is to do hypothesis testing in the period I and II for each parameter estimation obtained from the MLE method. The hypotheses used are:

- For $\mu$ parameter
  $H_0 : \mu = \mu_0$
  $H_1 : \mu \neq \mu_0$
  Test statistic: $Z_{hitung} = \frac{\hat{\mu} - \mu_0}{SE(\hat{\mu})}$

- For $\sigma$ parameter
  $H_0 : \sigma = \sigma_0$
  $H_1 : \sigma \neq \sigma_0$
  Test statistic: $Z_{hitung} = \frac{\hat{\sigma} - \sigma_0}{SE(\hat{\sigma})}$

- For $\xi$ parameter
  $H_0 : \xi = \xi_0$
  $H_1 : \xi \neq \xi_0$
  Test statistic: $Z_{hitung} = \frac{\hat{\xi} - \xi_0}{SE(\hat{\xi})}$

For the third hypothesis testing above, the decision reject $H_0$ if the value $|Z_{hitung}| > Z_{\alpha/2}$

### Table 2. Hypothesis Testing with the Normal Approach

| Period | $\mu$ Parameter | $\sigma$ Parameter | $\xi$ Parameter |
|--------|-----------------|-------------------|----------------|
|        | $\hat{\mu}$    | $\hat{\sigma}$   | $\hat{\xi}$   |
|        | $\mu_0$         | $\sigma_0$        | $\xi_0$        |
|        | $Z_{hitung}$    | $Z_{hitung}$      | $Z_{hitung}$   |
|        | $Z_{\alpha/2}$ | $Z_{\alpha/2}$    | $Z_{\alpha/2}$ |
|        | Decision        | Decision          | Decision       |
| 1      | 96              | 63                | 60             |
| (SE)   | 6,6             | 5,1               | 1,96           |
|        | 0               | Accept $H_0$      | Accept $H_0$   |
| 2      | 96              | 66                | 64,5           |
| (SE)   | 10,4            | 7,8               | -0,29          |
|        | 0               | Accept $H_0$      | Accept $H_0$   |

### Table 2 (Continued). Hypothesis Testing with the Normal Approach

| Period | $\hat{\xi}$ Parameter | $Z_{hitung}$ | $Z_{\alpha/2}$ | Decision |
|--------|------------------------|--------------|----------------|----------|
| 1      | $\xi_0$                | 0,1          | 0,1            | 1,96     | Accept $H_0$ |
| (SE)   |                        | 0,1          | 0,1            | 1,96     | Accept $H_0$ |
| 2      | $\xi_0$                | 0,1          | 0,1            | 1,96     | Accept $H_0$ |
| (SE)   |                        | 0,1          | 0,1            | 1,96     | Accept $H_0$ |
Based on Table 2 shows that the value of $|Z_{hitung}|$ for each parameter estimation $\mu, \sigma$, and $\xi$ in period I and II is less than the value of $Z_{\alpha/2}$ so the decision is accept $H_0$ and the meaning are $\mu = \mu_0$, $\sigma = \sigma_0$ and $\xi = \xi_0$. Because $\xi = \xi_0$ with $\xi_0 = 0$, so can defined that rainfall extreme data in Indramayu station with period I and II have Gumbel distribution. Therefore, there is no change in the distribution for period I and II and then it can be said that there is no climate change in Indramayu station with January 1960-December 2003 period.

| Reperiod       | Period I (1961-1990) | Period II (1991-2003) |
|----------------|----------------------|-----------------------|
|                | Time                 | Return Level          | Time             | Return Level          |
| 2 block = 6 months | January 1991-       | 120 mm                | January 2004-    | 120 mm                |
|                | Juni 1991            |                       | Juni 2004        |                       |
| 3 block = 9 months | January 1991-       | 157 mm                | January 2004-    | 157 mm                |
|                | September 1991       |                       | September 2004   |                       |
| 4 block = 12 months | January 1991-       | 182 mm                | January 2004-    | 182 mm                |
|                | Desember 1991        |                       | Desember 2004    |                       |

Based on Table 3 to see that maximum rainfall expected on average can be exceeded once in a period of 6, 9, and 12 months for period I was 120 mm; 157 mm; and 182 mm, while for the period II was 120 mm, 157 mm, and 182 mm.

7. Conclusion
Based on the results and discussion of the obtained some of the following conclusion:
There is no climate change in Indramayu station from January 1961-December 2003. This is shown by the absence of a change in distribution in period I (1961-1990) and period II (1991-2003) and the value of each parameter estimation obtained by the MLE method is lies in confidence interval of 95%.

8. References
[1]  Diebold F X, Schuermann T and Stroughair J D 1998 J. Risk Finance 1 30-36
[2]  Gilli and Kellezi 2003 An Application of Extreme Value Theory for Measuring Risk (Switzerland: Geneva)
[3]  Hastaryta R and Effendie A R 2006 Estimasi value-at-risk dengan pendekatan extreme value theory generalized pareto distribution (Studikasus IHSG 1997-2004)
[4]  Gander J P 2009 J. Economics 3 1-41
[5]  Georgescu V 2009 Measuring risk with extreme value theory Thesis CASE (Center of Applied Statistics and Economics) Humboldt University
[6]  Gourier E, Farkas W and Abbate D 2009 J. Operational Risk 4 3 3-26
[7]  Alves M L F and Gomes M L 1996 Statistical choice of extreme value domains of attraction – a comparative analysis Communication in statistics-theory and method 25 789-811
[8] Deane J H B, Johnstone G G and Ledford A W 1997 J. Telecommunication 33 832-833

[9] Kang J, Hamilton B A and McLean 2000 J. Telecommunication

[10] Kalyani S and Giridhar K 2006 J. Telecommunication and Computer Networks 6 2893-2898

[11] Hosking J R M, Wood E F and Wallis J R 1984 J. Technometrics 27 251-261

[12] Coles S, Heffernan J and Tawn J 1999 J. Applied Sciences 2 339-365

[13] Arshad M, Rasool M T and Ahmad I 2002 J. Applied Sciences 2 774-776

[14] Katz R W, Parlange M B and Naveau P 2002 Advances in Water Resources 25 1287–1304

[15] Klein T A M G and Können G P 2003 J. Climate 16 3665–3680

[16] Li Y, Cai W and Campbell E P 2005 J. Climate 18 852-863

[17] Gilleland E and Katz R W 2006 Analyzing seasonal to interannual extreme weather and climate variability with the extremes Toolkit (extRemes) Based on the Recommendation of the 18th Conference on Climate Variability and Change, 86th American Meteorological Society (AMS) Annual Meeting Atlanta-Georgia 29 January–2 February 2006

[18] Prang J D 2006 SebarannilaiekstremterampatdalamfenomenacurahhujanTesisJurusanStatistika IPB-Bogor

[19] Diebolt J, Guillou A and Naveau P 2008 J. Statistical 6 33-50

[20] Giles D E, Feng H and Godwin R T 2011 Econometrics Working Paper 35 200-207

[21] McNeil A J 1999 Extreme Value Theory for Risk Managers Department Mathematic ETH Zentrum-Zurich