Study of Daily Rainfall Distribution for Flood Disaster Mitigation in Bandar Lampung

Achmad Raflie Pahlevi[1,2], Warsono[1], Khorin Nisa[1], and Mustofa Usman[1]

[1]Faculty of Mathematics and Natural Sciences, University of Lampung
[2]Indonesia Agency for Meteorology Climatology and Geophysiscs (BMKG)

E-mail: achmad.raflie@bmkg.go.id

Abstract. Extreme rainfall is a hydro-meteorological event that most often causes disasters such as floods and landslides in Bandar Lampung. This makes the importance of using rainfall probability distribution to explain the potential of extreme rainfall events in Bandar Lampung. In this study, rainfall data were modelled using generalized Pareto distribution. The generalized Pareto distribution is known to be effective in explaining extreme event data and is suitable for data that involves time-dependent parameters to account for temporal changes in the frequency of distribution. The data used in this study are the intensities of daily rainfall from the Maritime Station of Meteorology, Climatology and Geophysics Agency (Indonesian: Badan Meteorologi, Klimatologi, dan Geofisika, abbreviated BMKG) in Panjang - Bandar Lampung in the period 1999-2018. The results showed that the generalized Pareto distribution was very suitable in describing the intensity of rainfall in Bandar Lampung and could be used for flood disaster mitigation.

Keyword: Extreme rainfall, generalized Pareto distribution, flood disaster mitigation

1. Introduction

Extreme rain event is one of the most extreme weather events in hydrometeorology, it gets more attention because of its large impact on social economy and human life. This makes the importance of using rainfall probability distribution to explain the potential of extreme rainfall events in Bandar Lampung [1]. Estimating a probability distribution model that fits for the intensity of daily rainfall and wind speed has long been an interesting research topic in hydrology and meteorology.

The use of extreme value distributions in hydrometeorology was first introduced by Jenkinson in 1955 [2]. Probability models have been applied successfully in many physical phenomena such as wind speed, rainfall, and air quality [3]. Normal, lognormal, Pearson, log-Pearson, exponential, Gumbel, generalized extreme value, Weibull, and generalized Pareto distribution are the most frequently used in hydrometeorology [4].

Generalized Pareto (GP) distribution was first introduced by Pickands in 1975 [5] and it becomes a stable distribution when its values above the threshold. The use of GP distribution has been applied many times in extreme value analysis of meteorological variables, such as rainfall [6], wind speed [7], and drought [8]. The probability density function (PDF) of GP distribution is as follows,

$$f(x) = \frac{1}{\sigma}\left(1 + \xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}-1}, \xi \neq 0 \quad (1)$$

And cumulative distribution function (CDF) of GP is

$$F(x) = 1 - \left(1 + \xi \frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}, \xi \neq 0 \qquad (2)$$

x in this paper is defined as the intensity of daily rainfall, ξ is shape parameter, σ is scale parameter, and μ is location parameter.

The purpose of this study is to show that GP distribution is the best fit distribution to describe the intensity of daily rainfall in Bandar Lampung. The use of GP distribution models in the probability of extreme rain events can be applied to flood disaster mitigation in Bandar Lampung.

2. Methodology

The data used in this study are daily rainfall intensity for 21 years from 1999 – 2018 obtained from Lampung Maritime Meteorological Station of BMKG in Bandar Lampung. The rainfall data are divided into four different seasons, that are the rainy season (December, January and February), transition season I (March, April and May), dry season (June, July and August), and transition season II (September, October and November).

Parameter estimation for GP distribution is done using the maximum likelihood estimation (MLE) method. MLE has been widely used, because it produces an efficient and consistent estimator. The estimation of GP-2 parameter distribution parameters with MLE was first carried out by Grimshaw (1993) [9] and continued being used by Coles [10], Chaouche [11], and Husler [12]. The MLE solution for θ is obtained by maximizing the likelihood function L, i.e. :

$$\frac{d \log L}{d\theta} = 0 \qquad (3)$$

To show that the GP distribution model is the best fit distribution model, the GP distribution model is compared to other distributions namely generalized extreme value (GEV) and Weibull. Distribution model of rainfall data is then verified by goodness of fit tests, here we use the Kolmogrov-Smirnov (KS) and Anderson-Darling (AD) tests. The smallest value shows the best distribution. With N is the amount of data, the KS test statistic is given by

$$D = \max_{1 \le i \le n} \left(F(x_i) - \frac{i-1}{N}, \frac{i}{N} - F(x_i) \right) \qquad (4)$$

and AD test statistic is given by

$$A^2 = -N - S \qquad (5)$$

where

$$S = \sum_{i=1}^{n} \frac{(2i-1)}{N} \left[\ln F(x_i) + \ln\left(1 - F(x_{N+1-i})\right) \right] \qquad (6)$$

The probability for the intensity of daily rainfall which is classified as extreme and passing a threshold value is explained as follows,

$$P(X > x) = 1 - P(X < x) = 1 - \int_{-\infty}^{x} f(x)dx \qquad (7)$$

3. Result and Discussion

Intensity of daily rainfall frequency data in Figure 1 shows that the information on intensity of rainfall data is positively skewed. It strengthens the reason for using extreme value distribution (such as GP, GEV, and Weibull). It shows that intensity of daily rainfall data the highest frequency of rainfall occurs in the rainy season, then transition 1 and transition 2, and the lowest frequency of rainfall during the dry season.

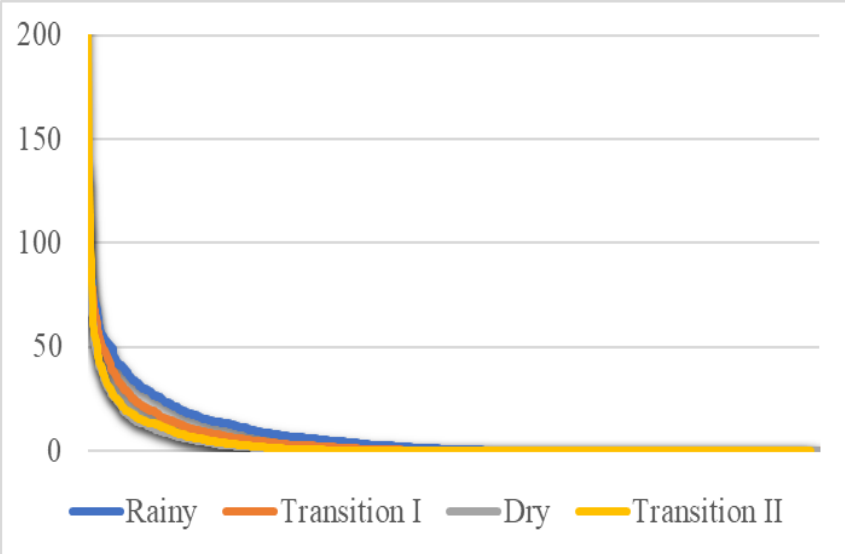

Figure 1. Graph of daily rainfall intensity data frequency

The data are summarized in Table 1 giving the minimum, maximum, median, mean, standard deviation, skewness and kurtosis. The summary in the table shows that even the highest rainfall frequency in the rainy season, the intensity of daily maximum rainfall in rainy is actually the lowest. The maximum of daily rainfall intensity occurs in transition II with rainfall intensity reach 204.9 mm. On transition II, larger value of skewness and kurtosis showed that the distribution of data was stretched to positively skewed, while lower value of skewness and kurtosis that data distribution is more integrated. It shows that during rainy season the data is more integrated while in the transition II season the data is more stretched to positively skewed.

Table 1. Summary of daily rainfall intensity data

	Min	Max	Median	Mean	Standard Deviation	Skewness	Kurtosis
Rainy	0	108.1	1	7.7814	14.5761	2.9336	13.2558
Transition I	0	130	0	5.3451	13.0916	4.3556	27.9305
Dry	0	137.2	0	3.2385	9.7559	5.3222	42.2808
Transition II	0	204.9	0	3.8118	11.9812	7.1529	79.1506

3.1. Goodness of Fit Test

In an attempt to prove that GP distribution is the best probability model distribution to describe daily rainfall data in Bandar Lampung, the goodness of fit test is carried out using KS and AD tests. The results of the KS and AD tests are shown in Table 2. Based on Table 2, GP distribution is the best distribution, because it has the lowest test statistic value. GP distribution shows very good compatibility for the rainy season and decreases when entering the dry season.

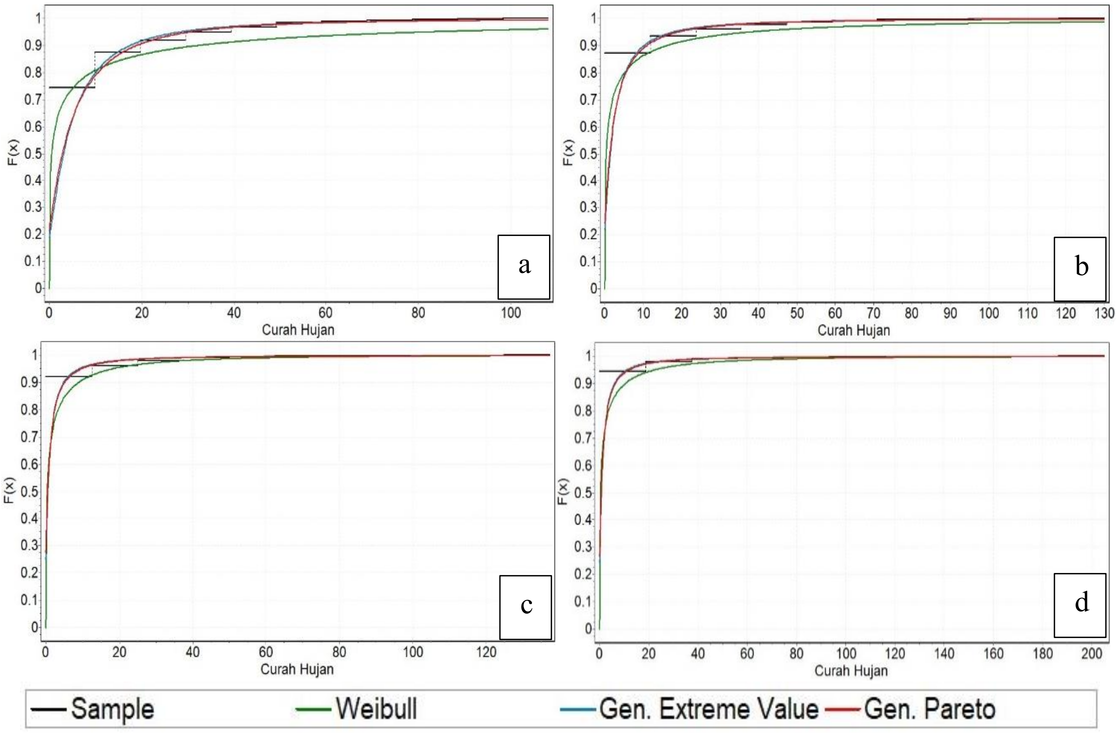

Figure 2. Graph showing the cumulative distribution function of the fitted distribution in all seasons: a) rainy, b) transition II, c) dry, and d) transition II

The cumulative distribution function (CFD) graph of the three distributions is shown in Figure 2. The graph shows how the GP, GEV and weibull distributions fit the data. In Figure 2, it can be seen that the GP distribution is the best distribution because it is closest distribution to the CDF of daily rainfall intensity data.

Tabel 2. Goodness of fit statistics

Distribution	Rainy		Transition II		Dry		Transition II	
	Kolmogorov - Smirnov	Anderson - Darling	Kolmogorov - Smirnov	Anderson - Darling	Kolmogorov - Smirnov	Anderson - Darling	Kolmogorov - Smirnov	Anderson - Darling
Generalized Pareto	0.21282	112.52	0.2901	178.83	0.40964	314.59	0.37153	263.37
Generalized Extreme Value	0.22574	132.76	0.31495	201.41	0.43474	347.59	0.39644	291.77
Weibull	0.4145	1117.2	0.53257	1580.1	0.68371	2272.6	0.63728	2017

3.2. Probability of daily rainfall intensity

GP distribution more suitable than other distributions for daily rainfall intensity data in Bandar Lampung. The probability calculation will use CDF from the GP distribution. The parameters of the GP distribution are shown in Table 3.

Table 3. Parameter estimates of GP distribution

	ξ	σ	μ
Rainy	0.46673	4.8671	-1.2323
Transition I	0.6373	2.185	-0.66375
Dry	0.75616	0.85725	-0.31068
Transition II	0.73615	1.0992	-0.38128

Based on Table 3, the CDF of rainy season,

$$F(x) = 1 - \left(1 + 0.46673\frac{x + 1.2323}{4.8671}\right)^{-0.46673} \quad (8.a)$$

CDF of transition I season,

$$F(x) = 1 - \left(1 + 0.6373\frac{x + 0.66375}{2.185}\right)^{-0.6373} \quad (8.b)$$

CDF of dry season,

$$F(x) = 1 - \left(1 + 0.75616\frac{x + 0.31068}{0.85725}\right)^{-0.75616} \quad (8.c)$$

CDF of transition II season,

$$F(x) = 1 - \left(1 + 0.73615\frac{x + 0.38128}{1.0992}\right)^{-0.73615} \quad (8.d)$$

and x is intensity of daily rainfall data.

Table 4. Probability of daily rainfall intensity (in percentage)

x	> 0	≥ 25	≥ 50	≥ 100
Rainy	79	6.8	2.2	0.6
Transition I	76	3.5	1.3	0.4
Dry	73	1.5	0.6	0.2
Transition II	73	2	0.8	0.3

By using CDF parameters from Table 3 and equation (8.a) - (8.d), the probability of daily rainfall intensity data is obtained in Table 4. During the rainy season, the probability of rain with an intensity above 100 mm or very heavy rainfall intensity is,

$$P\left(X_{rainy} > 100\right) = 1 - P\left(X_{rainy} \leq 100\right) = 1 - 0.994 = 0.006$$
$$P\left(X_{rainy} > 100\right) = 0.6\%$$

when transition I season,

$$P(X_{transition\ II} > 100) = 1 - P(X_{transition\ I} \leq 100) = 1 - 0.996 = 0.004$$
$$P(X_{transition\ I} > 100) = 0.4\%$$

when dry season,

$$P\left(X_{dry} > 100\right) = 1 - P\left(X_{dry} \leq 100\right) = 1 - 0.998 = 0.002$$
$$PX_{dry} = 0.2\%$$

when transition II season,

$$P(X_{transition\ II} > 100) = 1 - P(X_{transition\ II} \leq 100) = 1 - 0.997 = 0.003$$
$$P(X_{transition\ II} > 100) = 0.3\%$$

So that, the frequency of very heavy rainfall from the biggest to the lowest in a row, is the rainy season (0.6%), transition period I (0.4%), dry season (0.2%), and transition II (0.3%).

4. Conclusion

In this paper, we have shown that the intensity of daily rainfall data is positively skewed. Based on goodness of fit tests, GP distribution is the best fit distribution to describe daily rainfall intensity data. The frequency of daily rainfall intensity that more than 100 mm or very heavy rainfall in a row, is the rainy season (0.6%), transition I (0.4%), dry season (0.2%), and transition II (0.3%).

5. Acknowledgment

Financial support for this study was provided through an operating grant from the University of Lampung. Daily rainfall intensity data was provided by Lampung Maritime Meteorology Station, Indonesia Agency for Meteorology Climatology and Geophysics (BMKG).

6. References

[1] Pahlevi, A.C., dan Warsono. 2018. Kajian Best-Fit Distribusi Probabilitas Untuk Curah Hujan Harian dan Aplikasinya Dalam Mitigasi Hujan Ekstrim Di Pulau Sumatera. Dipersentasikan di: Seminar Nasional Metode Kunatitatif II. Tersedia: http://matematika.fmipa.unila.ac.id/wp-content/uploads/2019/04/Prosiding-Seminar-Nasional-Metote-Kuantitatif-2018.pdf

[2] Jenkinson, A.F. 1955. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society Vol 81: 158-171*

[3] Oguntunde, P.E., Odetunmibi, O.A., dan Adejumo, A.O. A study of probability models in monitoring enviromental pollution in Nigeria. *Journal of Probability and Statistics.Doi: 10.1155. 2014*

[4] Alam, M.A., Emura, K., Farnham, C., Yuan, J. 2018. Best-Fit Probability Distributions and Return Periods for Maximum Monthly Rainfall in Bangladesh. *Journal of Climate 6 (9); doi:10.3390*

[5] Pickands, J. 1975. Statistical Inferences Using Extreme Order Statistics. The Annlas of Statistics, Vol.3, No.1

[6] Acero, F.J., Garcia, J.A., dan Gallego, M.C. 2011. Peaks over Threshold Study of Trends In Extreme Rainfall over the Iberian Peninsula. *Journal of Climate. DOI: 10.1175]*

[7] Simiu, E. 2006. Generalized Pareto methods for wind extremes. Useful tool or mathematical mirage? Journal of Wind Engineering and Industrial Aerodynamics Vol. 95 pp 133-136

[8] Nadarajah, S. Generalized Pareto models with application to drought data. *Environmetrics Vol. 19:395-408*

[9] Grimshaw, Scott. (1993). Computing Maximum Likelihood Estimates for the Generalized Pareto Distribution. *Technometrics. 35. 185-191. 10.1080/00401706.1993.10485040.*

[10] Coles, S. 2001. *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics.

[11] Chaouche, A., dan Bacro, J.N. 2004. A statistical test procedure for the shape parameter of generalized pareto distribution. *Computational Statistics and Data Analysis Vol.45:787-803*

[12] Husler, J., Li, D., Raschke, M. 2011. Estimation for the Generalized Pareto Distribution Using Maximum Likelihood and Goodness of Fit. *Comunications in Statistics-Theory and Methods, Vol 40:2500:2510*