Extremity of Rainfall Distribution in Palembang

By Herlina Hanum Herlina Hanum

WORD COUNT	2910	TIME SUBMITTED	09-MAR-2020 05:31PM
		PAPER ID	56287715

Extremity of Rainfall Distribution in Palembang

Herlina Hanum, Novi Rustiana Dewi, Dwi Setyawan

Abstract —Extreme rainfall is one of focus research in extreme values study. Some probability distributions had been used to model extreme rainfall. This research studies the extremity of Palembang's daily rainfall. The data set is 10 years ofdaily rainfall during January 2009-Desember 2018. It is only 10-year data due to the information from BMKG (Indonesian Bureau for meteorology, climatology, and Geophysics) that there is a change in the normal rainfall in the last ten years. In this research we study the distribution of whole rainfall, rainfall at least 5 mm/day, and extreme rainfall. Extreme rainfall data is obtained by POT (Peaks over threshold) of quantile 90 %. There are 181 rainfalls exceeding the threshold of 53 mm/day. The data distributions are fitted model by Exponential, Gamma, Weibull, and Gamma-Pareto. The data try to be judged as sub exponential and hyper-exponential according to the estimator of parameter. As the result, Weibull distribution has good fit to the whole rainfall data with MAPE 9.3% and P-value of Kolmogorov-Smirnov test 0.1316. The distribution of whole data belong to sub exponential distribution since the Weibul's shape parameter $\alpha < 1$. The rainfall data with values at least 5mm/day were good estimated by Exponential, Gamma, Weibull, and Gamma-Pareto according to MAPE. On the other hand, Kolmogorov-Smirnov test notes that these distributions are not appropriate with the distribution of the data. Gamma and Gamma-Pareto distribution is very good to fit extreme rainfall in Palembang. The extreme rainfall has skewness of Gamma-Pareto more than 16 and tend to be sub exponential according to Gamma distribution.

Index Terms— Extreme rainfall, Gamma, exponential, Weibull, gamma-pareto, sub exponential, Palembang

———————————— ◆ ————————————

1INTRODUCTION

EXTREME rainfall is one of focus research in extreme values study. Some probability distribution had been used to model extreme rainfall. Extreme rainfalls are rainfalls which intensity more than its normal value. The rainfall may be daily, ten-days, or monthly. According to BMKG[1], the extreme daily rainfall is rainfall more than 100 mm in 24 hours. Some study noted that extreme values are which more than a threshold. This method is known as Peaks over threshold (POT) [2], [3]. Chavez-Demoulin et al. [4] used threshold of 90% quantile of the data set.

Some probability distributions have been used to model the extreme values; these are Pareto, Gamma and Weibull see [5],[6],[7]. According to the tail of probability distributions, compares to the tail of exponential distribution, they are divided into two groups, i.e. hyper-exponential and sub-exponential [8]. Papalexiou et al. [8], sub exponential is distribution with CDF $F(x)$ which fulfill (1)

$$lim_{x\to\infty}\frac{1-F(x)}{exp\left(-x/_{\beta}\right)}=\infty\ ,\forall\beta>0 \quad (1).$$

Meanwhile, according to Rezaul and Grout [9] heavy tail distribution follows

$$lim_{x\to\infty}\frac{d(1-F(x))}{d\log(x)}=-\alpha,\quad 0<\alpha<2 \quad (2)$$

Sub-exponential or heavy tail or long tail distribution is used to model extreme values. Gamma and Weibull with shape parameter β >1 belong to hyper-exponential, while Pareto and Weibull with shape parameter $\beta < 1$ belong to sub-exponential distribution. Papalexiou et al. [8] note that Weibull and Pareto better fit to extreme daily rainfall distribution than Gamma does.

Gamma-Pareto (G-P) distribution is join distribution between Gamma and Pareto. Hanum et al. [10] notes that G-P distribution has very good fit to some extreme rainfall data distribution. It is better fitted to the data than Pareto and Gamma distribution. G-P is not yet determined as hyper-exponential or sub exponential distribution. Alzaatreh et al.[11] note that the skewness of the distribution increased as the shape parameter c increased. The pdf of G-P is (3)

$$g(y)=\frac{\theta^{-1}}{c^{\alpha}\Gamma(\alpha)}\left(\log\left(\frac{y}{\theta}\right)\right)^{\alpha-1}\left(\frac{y}{\theta}\right)^{\left(-1/_{\varrho}-1\right)} \quad (3)$$

With $\alpha, c, \theta > 0$ and $x > \theta$. Parameter $c = \beta/k$, where β the scale parameter of Gamma and k the shape parameter of Pareto. In this research, we also study the properties of this distribution in term of hyper and sub exponential.

Palembang city, as the capital of South Sumatra Province Indonesia, is located in tropical country. It has many rainy days each year. The question is what type of distribution that better fits to the rainfall in Palembang. Which distribution better fits to rainfall exceed than zero, and which distribution better fits to the extreme rainfall.

2 DATA AND METHODS

The data used in the study are daily rainfall observed at Sultan Mahmud Badaruddin Station from January 2009 to December 2018. They were downloaded from https://www.ogimet.com. There are 3258 observed days. The extreme rainfalls are determined via POT of 90% quantile.

There are four distributions used to model the rainfall distribution. They are Exponential (β), Gamma(α, β), Weibull(α, β), and Gamma-Pareto (θ, α, c). The pdf and cdf of Exponential, Gamma, and Weibull, respectively, are

————————————

- *Herlina Hanum is currently Associate Professor in Statistics in University of Sriwijaya, Indonesia, PH-+6281367732597. E-mail: linhanum@gmail.com*
- *Novi Rustiana Dewi is currrently Lecturer in Mathematics in University of Sriwijaya, Indonesia, PH-+62711580765. E-mail: novi_rustianadewi@yahoo.co.id*
- *Dwi Setyawan is currently Associate Professor in Soil and Environmental Sciences in University of Sriwijaya, Indonesia, PH +62711580059. E-mail: dsetyawan@unsri.ac.id*

Exponential(β) distribution :

$$f_E(x) = \left({}^{x}/_{\beta}\right) exp\left(-{}^{x}/_{\beta}\right),$$

$$F_E(x) = 1 - exp\left(-{}^{x}/_{\beta}\right),\ x \geq 0\ , \beta > 0$$

Gamma(α, β) distribution

$$f_G(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} exp\left(-\frac{x}{\beta}\right)$$

$$F_G(x) = 1 - \Gamma\left(\alpha, \frac{x}{\beta}\right)/\Gamma(\alpha), \qquad 0 < x < \infty$$

where $\alpha, \beta > 0, \Gamma(\alpha) > 0,$ and $\Gamma\left(\alpha, \frac{x}{\beta}\right) = \int_{x/\beta}^{\infty} e^{-z} z^{(\alpha-1)} dz$ the upper incomplete gamma function [12].

Weibull(α, β) distribution in [8]

$$f_W(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} \ exp\left(-\left(\frac{x}{\beta}\right)^{\alpha}\right)$$

$$F_W(x) = 1 - exp\left(-\left(\frac{x}{\beta}\right)^{\alpha}\right), 0 < x < \infty$$

where the shape parameter $\alpha > 0$ and scale parameter $\beta > 0$.

We used Akaike's Information Criteria (AIC)[13] and Mean Absolute Percentile Error (MAPE)[14] as measurement of the goodness of fit. We also test the goodness of fit of the distribution with Kolmogorov-Smirnov test [15].

3 RESULTS AND DISCUSSION

The result in this paper begins with the study of the tail of G-P to determine whetherit is a sub exponential or not. Modeling daily rainfall for rainy days is the next discussion. It is followed by modeling the extreme rainfalls.

3.1 The tail of Gamma-Pareto

Here we want to know if G-P distribution belongs to hyper-exponential or sub exponential distribution. We use the criteria (1) that certain distribution belongs to sub exponential distribution if it follows (1). The CDF of G-P is

$f(x) = \gamma(\alpha, z)/\Gamma(\alpha)$

where $z = c^{-1}\log(x/\theta)$ and $\gamma(\alpha, z) = \int_0^{v} e^{-z} z^{(\alpha-1)} dz$ the lower incomplete gamma function, with $\theta, \alpha, c, > 0$ and $x > \theta$.Hence the exceedence probability function (EPF) of G-P is (3)

$$\bar{F}(x) = \Gamma(\alpha, z)/\Gamma(\alpha) \qquad (4).$$

The EPF of G-P similar with the EPF of Gamma, that is both of them are in the form of upper incomplete gamma divided by complete gamma function. The EPF of Gamma (α, β) is (4)

$$\bar{F}(x) = \Gamma(\alpha, x/\beta)/\Gamma(\alpha) \qquad (5).$$

The limit form of the ratio of Gamma and exponential EPF in [8] is 0 when $\beta_G < \beta_E$ and ∞ when $\beta_G > \beta_E$. If $\beta_G = \beta_E$ the limit is (6)

$$\lim_{x\to\infty} \frac{\bar{F}(x)}{\exp\left(-{}^{x}/_{\beta_E}\right)} = \begin{cases} 0, & 0 < \alpha < 1 \\ 1, & \alpha = 1 \\ \infty, & \alpha > 1 \end{cases} \qquad (6)$$

Where β_G and β_E are respectively β for Gamma and Exponential. Since the parameter G-P, $c = \beta/k$, then $z = (k/\beta)\log(x/\theta)$. Therefore, the behavior of G-P in the limit term, with respect to β, will be similar to the behavior of Gamma, although z go to ∞slower than x/β as $x \to \infty$.

3.2 Modeling whole rainfall data

Palembang had 1801 rainy daysout of 3285 observed days. It means Palembang only had 181rainy days per year in average. Furthermore, only 1222 days received rainfall at least 5 mm/day. So more than half - year Palembang has dry and almost dry days. The whole data means the rainfall of 1801 rainy days. This data set contains 579 or 1/3 values in interval 0.1 - 4.9 mm/days.

TABLE 1
PARAMETERS ESTIMATION FOR WHOLE DATA

Distribution	Parameter	AIC	MAPE	P-value
Exponential	0.0488	14484	0.4443	7.416e-14
Gamma	0.7018 0.0342	14314	0.1498	0.0148
Weibull	0.7815 17.6293	14271	0.09338	0.1316
G-P	0.1 6.9542 0.6417	14610	0.5838	4.264e-12

It is rather smaller compared with 1801 rainy days in 10 years. On the other hand, it has 50 values more than 100mm/day (the extreme rainfall according to BMKG) including 3 highest values of 228.5, 264, 429 mm/day. The histogram of whole data will be very high at lower values and very long tail at the right.

Table 1 contains parameter estimation of four distributions and their goodness of fit. According to the value of AIC, the order of best to worst estimator is Weibull, Gamma, Exponential, and G-P. With MAPE less than 10 %, Weibull with α< 1 has very good fit to the data [14]. While Gamma with $\alpha < 1$ gives a good estimation with MAPE 10-20%. Furthermore, the result of K-S test note that only Weibull which is not significant at 5% significant level. It means that this data set is only appropriate for Weibull [15]. Since α< 1 for Weibull, this data set belongs to sub exponential distribution [8].

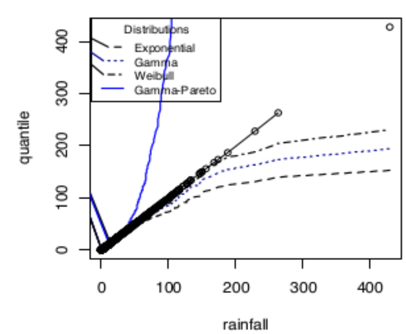

Fig. 1.Quantile plot for whole data.

Fig. 1 shows the estimation of whole data by the four distributions. Three distributionsunder estimate the extremes values of rainfall. Weibull could be very good model the distribution of whole data except for 3 highest values of daily rainfall. On the other hand, G-P is the only distribution that over estimate the rainfall. Hanum et al. [4] noted that G-P is not good to estimate the data with a lot of ties at low values. G-P distribution will underestimate the low values and over estimate the rest. The large MAPE values of G-P mostly come from the low values.

3.3 Estimation for rainfall at least 5 mm/day

This data set was analyzed to study the distribution of rainfall without 1/3 part of the data which values less than 5 mm/day. According to MAPE values, in Table 2, that fall in range 10- 20 %, all four distributions give good estimate for this data. Here G-P best fit the data with smallest AIC and MAPE, and largest K-S test P-value. Although MAPE values indicated that none of the distributions is suitable for the data at 5 % significant level. Only G-P is not significant at 1% level. Unfortunately, it is not good enough for fitting datasince, for K-S test, the larger p-value, the better fit.

TABLE 2
PARAMETERS ESTIMATION FOR WHOLE DATA
WITHOUT VALUES LESS THAN 5MM/DAY

Distribution	Parameter	AIC	MAPE	P-value
Exponential	0.0342	10698.15	0.1935	1.584e-13
Gamma	1.4703 0.0502	10600.97	0.1818	1.034e-06
Weibull	1.1439 30.9434	10659.83	0.1929	1.364e-07
G-P	5 2.3001 0.6190	10090.97	0.1438	0.01563

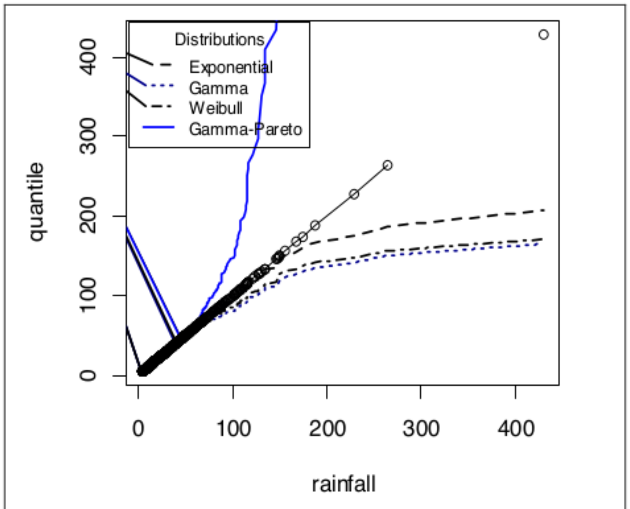

Fig. 2. Quantile plot for rainfall at least 5mm/day.

There is an ambiguous of sub and hyper exponential distribution of the data. According to values of β for Exponential and Gamma, that is $\beta_G > \beta_E$, the distribution of the data is sub exponential. On the other hand, Weibull with α >1, indicates the data set is hyper-exponential. Since parameter c for G-P is almost similar with that for whole data and also with α >1, it might be said that the data are still sub exponential.

In Fig. 2 can be seen that the behavior of quantile function of four distributions is similar with their estimation for the whole data. The different is the Exponential distribution could give better estimate for the higher values except for 6 highest values. On the other hand, the quantile function of G-P shows that the portion of the data being overestimate is smaller than that of for whole data. It also can be seen that the estimation of Gamma is closer to the estimation of Weibull and both are not good estimation for higher values.

3.4. Model for Extreme Rainfall

This data set is rainfall which values exceed the 90 % quantile of the data. The 90% quantile is 53 mm/day. This value is closed to the threshold of heavy rain, that is 50 mm/day, used by BMKG. There are 181 values which exceed the threshold

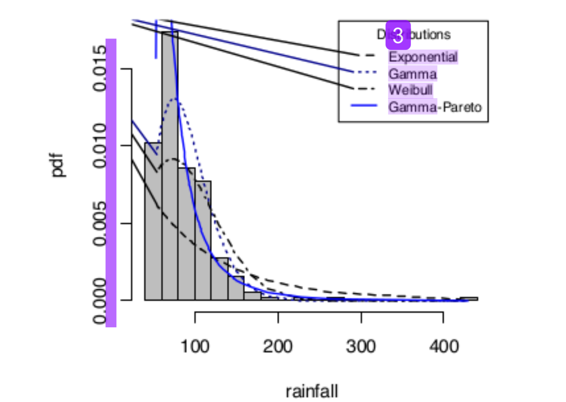

Fig. 3. Histogram of extreme values and pdf of the distributions.

In Fig 3 the histogram of extreme data has the modal that apart from the minimum value. The distribution of the data has shorter tail than the distribution of the two previous rainfall data. We can see that all three distributions have tail thinner than the tail of exponential. It means they are hyper-exponential. Since the data better fit to Gamma and G-P, consequently, according to the tail, it could be noted that the distribution of the extreme rainfall is not a heavy tail distribution. In further look, the pdf of G-P become fatter then tail or at least similar of Exponential rainfall more than 400 mm/day. So the conclusion might be different for G-P.

Fig. 4 shows how G-P estimates the extreme data. It is underestimate the lower and the highest values and overestimate the middle values. It could approach all values include the tail of the distribution of extreme rainfall data. Meanwhile, Gamma and Weibull estimation run away from highest values.

This extreme data set is better fitted to Gamma and G-P. Both have smaller AIC. With MAPE less than 10 %, as can be seen in Table 3, both distributions have very good estimate for the data. G-P distribution very good fit the data with smallest AIC, MAPE only 3.7 %, and P-value close to 1.

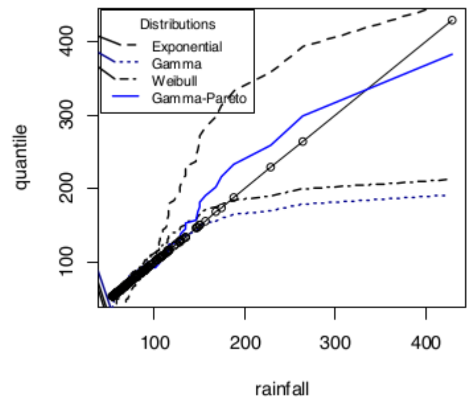

Fig. 4. Quantile plot for estimation of extreme rainfall

TABLE 3
PARAMETERS ESTIMATION FOR EXTREME DATA

Distribution	Parameter	AIC	MAPE	P-value
Exponential	0.0115	1981.48	0.4362	2.2e-16
Gamma	7.2141 0.0827	1760.16	0.0879	0.08297
Weibull	2.1411 98.2507	1829.67	0.1682	0.0001
G-P	53 1.2616 0.3403	1637.84	0.0374	0.8918

The G-P distribution with α =1.2616 and c = 0.3403 has skewness more than 16.443. It has truly right skew distribution. The question is that it is skew enough to be a heavy tail distribution. In Table 3 Gamma distribution, which very good fit the data, has $\beta_G > \beta_E$ to indicate the data follows the distribution of sub exponential. This conclusion is different from the pdf plot, even though they both come from the comparison of other distribution with Exponential

4 CONCLUSION

Based on the parameters of four tested distributions again rainfall data in Palembang, we conclude that

1. The behavior of the tail of G-P follows that of Gamma
2. Weibull distribution good fits whole rainfall data which belong to sub exponential distribution.
3. The rainfall data with values at least 5mm/day are good estimated by Exponential, Gamma, Weibull, and Gamma-Pareto according to MAPE but they are not good enough to model the data.
4. Gamma-Pareto distribution has very good fit to extreme rainfall in Palembang.
5. The extreme rainfall has skewness of Gamma-Pareto more than 16 and tends to be sub exponential according to Gamma distribution.

Acknowledgments

This work was funded by Universitas Sriwijaya under Competitive Research grant (PNBP funding).

REFERENCES

[1] BMKG (Indonesia Bureau for Meteorology, Climatology, and Geophysics) Provinsi Banten dan DKI Jakarta, "Buletin Analisis dan Prakiraan Hujan,"edisi Juni 2018 (Bulletin Analysis and estimation of rainfall, June 2018)

[2] F. Mallor, E. Nualart, and G. Onner, An introduction to statistical modeling of extreme values. Application to calculate extreme wind speeds, Hogeschool Universteit Brussel Research Paper 2009/36, (2009), 45.

[3] Roth, M., T. A. Buishand, G. Jongbloed, A. M. G. Klein Tank, and J. H. van Zanten, "A regional peaks-over-threshold model in a nonstationary climate," *Water Resour. Res.*, 48, W11533, doi:10.1029/2012WR012214, 2012

[4] V. Chavez-Demoulin, P. Embrechts,and S. Sardy, "Extreme-quantile tracking for financial time series," *Journal of Econometrics*, vol. 181, Issue 1, pp. 44-52, July 2014

[5] S. Dan'azumi, S. Shamsudin, and A. Aris, "Modeling the distribution of rainfall intensity using hourly data," *Am. J. Environ. Sci.* 6(3):238–243, 2010,ISSN 1553-345X

[6] M. A. Sharma and J. B. Singh, "Use of probability distribution in rainfall analysis," *New York Science Journal*, vol. 3, no. 9, pp. 40-49, 2010

[7] G. J. Husak, J. Michaelsen, and C. Funk, "Use of the Gamma Distribution to represent monthly rainfall in Africa for drought monitoring applications," *International Journal* of *Climatology*, vol. 27, no.7, pp. 935-944, 2007, http://dx.doi.org/10.1002/joc.1441

[8] S. M. Papalexiou, D. Koutsoyiannis, and C. Makropoulos, "How extreme is extreme? An assessment of daily rainfall distribution tails,"*Hydrology and Earth System Sciences*, vol. 17, no. 2, pp. 851-862, 2013, http://dx.doi.org/10.5194/hess-17-851-2013

[9] K.M., Rezaul and V. Grout, "A Comparison of Methods for Estimating the Tail Index of Heavy-tailed Internet Traffic. In: Sobh T., Elleithy K., Mahmood A., and Karim M. (eds) Innovative Algorithms and Techniques in Automation, Industrial Electronics and Telecommunications. Springer, Dordrecht, 2007

[10] H. Hanum, A.H. Wigena, A. Djuraidah, and I.W. Mangku, "Fitting extreme rainfall with Gamma-Pareto distribution,"*Applied Mathematical Sciences*, vol.9, no.121, pp. 6029-6039, 2015, HIKARI Ltd, http://dx.doi.org/10.12988/ams.2015.57489.

[11] A. Alzaatreh, F. Famoye, and C. Lee, "Gamma-Pareto Distribution and application,"*J. of Modern Applied Statistical Methods*, vol. 11, no. 1, pp. 78-94, 2012

[12] R.V. Hogg, J. Mckean, and A.T. Craig, *Mathematical Statistics* 7th edition, 2013, Pearson Education International

[13] C. D. Lewis, "Industrial and Business Forecasting Methods,London: Butterworths, 1982, http://dx.doi.org/10.1002/for.3980020210

[14] K.P. Burnham and D.R. Anderson,"Multinomial Inference, understanding AIC and BIC in model selecton," *Sociological Methods & Research*, vol.33, no. 2, pp. 261-304, Nov. 2004, doi: 10.1177/0049124104268644/2004 Sage publication

[15] H. Hassani and E.S. Silva, "A Kolmogorov-Smirnov Based Test for Comparingthe Predictive Accuracy of Two Sets of Forecasts,"*Econometrics*,vol. 3, pp.590–609, 2015, https://doi.org/10.3390/econometrics3030590

Extremity of Rainfall Distribution in Palembang

ORIGINALITY REPORT

3%

SIMILARITY INDEX

PRIMARY SOURCES

1	www.m-hikari.com Internet	16 words — 1%
2	rmets.onlinelibrary.wiley.com Internet	9 words — < 1%
3	Rashid, M. Mamunur, Simon Beecham, and Rezaul K. Chowdhury. "Statistical characteristics of rainfall in the Onkaparinga catchment in South Australia", Journal of Water and Climate Change, 2015. Crossref	8 words — < 1%
4	Md. Habibur Rahman. "Fitting of Probability Distribution on the Post-Monsoon Rainfall of Different Locations in Bangladesh", Environment and Natural Resources Research, 2019 Crossref	8 words — < 1%
5	Okjeong Lee, Inkyeong Sim, Sangdan Kim. "Application of the non-stationary peak-over-threshold methods for deriving rainfall extremes from temperature projections", Journal of Hydrology, 2019 Crossref	8 words — < 1%
6	dias.library.tuc.gr Internet	8 words — < 1%
7	Herlina Hanum, Novi Rustiana Dewi, Dwi Setyawan. "Modeling Indonesia inflation data: gamma-normal or normal", Applied Mathematical Sciences, 2018 Crossref	8 words — < 1%

Herlina Hanum, Aji Hamim Wigena, Anik Djuraidah, I Wayan

8	Mangku. "Modeling extreme rainfall with Gamma-Pareto distribution", Applied Mathematical Sciences, 2015 Crossref	7 words — < 1%

EXCLUDE QUOTES	OFF	EXCLUDE MATCHES	OFF
EXCLUDE BIBLIOGRAPHY	ON		