Extreme And Distribution Data Of Drainage System In The City Of Ambon

Obednego Dominggus Nara[1], Muhammad Bisri[2], Aniek Masrevaniah[3], Agus Suharyanto [3]

[1]Student of Doctoral Program in Civil Engineering, Faculty of Engineering, Brawijaya University Malang
[2]Professor of Civil Engineering, Faculty of Engineering, Brawijaya University Malang
[3]Lecturer in the Department of Civil Engineering, Faculty of Engineering, Brawijaya University Malang (Indonesia)

ABSTRACT

Climate change has been a significant environmental issue during this century. Various studies have been conducted in order to identify the causes of climate change. To avoid the harmful effects of extreme rainfall events and climate, it is necessary to give special attention to the extreme values, given the inability of human beings to avoid or escape from disaster that is caused by phenomena of precipitation and climate (*temperature*). One theory that specifically addresses extreme events is EVT (*Extreme Value Theory*). The purpose of this study is to provide an overview of extreme, distribution and modeling values as well as diagnostics model which is modeled with GEV distribution. The family of GEV distribution has three sub-families namely Gumbel, Fréchet and Weibull. This distribution is characterized by three parameters: the location, scale, and shape parameter. From the data distribution test with error rate (α) of 0.05, it gives a result that the tested distribution fits with GEV and Normal distribution. For the suitability distribution, in addition to specific inferential test, it can also be tested using the P-P plot and Q-Q plot. While empirical CDF graph is used to evaluate the fit of the data compared to the data distribution of the cumulative distribution of samples in which both probability values gives a similar profile at both levels axis. Based on return level of 100 years from the analysis of the maximum temperature, it increases only 10.66% or the xk value = 29.367 °C which means that in the period (*return period*) of 100 years, temperatures in the study area of Ambon will surpass 29.367 °C. Therefore, it is certain that the temperature changes in the study area are not too significant and have extreme phenomenon compared to the one with the maximum rainfall.

Keywords - *Climate change, GEV distribution, return level*

I. INTRODUCTION

Climate change has been a significant environmental issue during this century. Various studies have been conducted in order to identify the causes of climate change. Various measurable changes and consequences are needed to response and adapt to climate change, particularly an adaptation that can be done in urban areas. By 2030, an estimated 60% of the world population will live in urban areas, this is a tough challenge. IPCC (*Intergovernmental Panel on Climate Change*) found that, over the last 100 years (1906-2005) the Earth's surface temperature has risen to an average of about 0.74 °C, with greater warming mainly occurred at the land rather than at the sea. The late 1990s and the early 21st century were the warmest years since the existence of modern data archive. Warming increase by 0.2 °C is projected to occur for each decade and of the next two decades [8]. The issue of global climate change impacts is deeply felt in several regions in Indonesia, which contributes to the occurrence of some extreme phenomena, and one of the affected areas is the city of Ambon (*Moluccas-Indonesia*). To avoid the harmful effects of extreme rainfall events and climate, it is important to give special attention against extreme values, given the inability of human beings to avoid or escape from disaster that is caused by rainfall and climate phenomena. One theory that specifically addresses extreme events is EVT (*Extreme Value Theory*) [2,13]. EVT focuses on the information of extreme events based on the extreme values obtained to form the distribution function from the extreme values of the anomalous precipitation and climate (*temperature*).

II. METHOD

Some types of extreme value, the distribution, analysis and diagnostics of the model will be a method in the concept of the maximum value of the random variable modeling. Based on the theory, asymptotically, the extreme values of rainfall and temperature distribution will converge following the GEV (*Generalized Extreme Value*) [1,5,9,10,12] which is an extreme value in a given period. Suppose we have independent random variables x1, x2, ... xn, each variable xi has the same distribution function F*(x)*, which later on will be

considered that the maximum Mn = max (X_1, X_2, X_n) as written as the following equation;
The ξ parameter determines the characteristics of the tip of the distribution: if ξ <0 then the probability function has a finite right tip point and if ξ ≥ 0 the probability function will have an infinite right end point of. Extreme value distribution, introduced by Jenkinson (1955), is a combination of three types of extreme limited value distributions to a single form as derived by Fisher and Tippet (Hosking et al. 1985) [4]. Those three types are Gumbel, Fréchet and Weibull distributions whose equation are as follow

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

$$G(x)=\exp\left\{-\exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right\}; -\infty<x<\infty$$

The ξ parameter determines the characteristics of the tip of the distribution: if ξ <0 then the probability function has a finite right tip point and if ξ ≥ 0 the probability function will have an infinite right end point of. Extreme value distribution, introduced by Jenkinson (1955), is a combination of three types of extreme limited value distributions to a single form as derived by Fisher and Tippet (Hosking et al. 1985) [4]. Those three types are Gumbel, Fréchet and Weibull distributions whose equation are as follows:

$$G(x)=\begin{cases}\exp\left\{-\left[-\left(\frac{x-\mu}{\sigma}\right)\right]^{a}\right\} & ;x<\mu\\ 1 & ;x\geq\mu\end{cases}$$

$$G(x)=\begin{cases}0 & ;x\leq\mu\\ \exp\left\{-\left(\frac{x-\mu}{\sigma}\right)^{-a}\right\} & ;x>\mu\end{cases}$$

μ is location parameter, σ> 0 is scale parameter ξ> 0 is shape parameter (Stephenson, 2003)[9]

2.2 Parameter Estimation

Estimation method that is frequently used is the maximum likelihood estimation (MLE) [3,12]. Suppose y1, ..., yn are n independent observations of random variables Y1, ..., Yn which are GEV distributed respectively with indeterminate parameter θ = (μ, σ, ξ). In addition, the data is assumed not dependent one another, then: F(*yi,...,yn*|*θ*)=f(*y1*|*θ*)...*f*(*yn*|*θ*). This is called the likelihood function, often referred to l(*θ*|*y*). Log of the likelihood function:

$$L(\theta|y)=K\prod_{i=1}^{n}\exp\left(-\left(1+\xi\frac{ti-\mu}{\sigma}\right)^{-1/\xi}\right)$$

$$\prod_{i=1}^{n}\frac{1}{\sigma}\left(1+\xi\frac{yi-\mu}{\sigma}\right)^{\frac{1}{\xi}-1}$$

2.3 Model Diagnosis

A data modeling usually begins with assumptions, the data is assumed that it adheres to a particular distribution. Subsequently, with the data and the parameters of the distribution to be estimated. Consequently, the initial assumption must be tested, that is the similarity of data with a model that assumed. Two diagnostic test techniques based on the graph will be explained as follows.

2.4 Probability and QQ Plots Graphs

Suppose *y1*, ..., *yn* is a series of observations of extreme values (block maxima) which are independent with function of G distribution (unknown). If yi is assumed as GEV distributed, and from MLE, we get the estimation of the G distribution. The accuracy in estimating G can be tested by comparing the estimation model to the empirical distribution Suppose *y(1), y(2), .., y(n)* is a data sequence such a way that *y(1)* ≤ *y(2)* ≤ ..... ≤ *y(n)*, then the empirical function is easily obtained, as follows,

$$P(Y\leq y_{(i)}=\tilde{G}(y_{(i)})=\frac{i}{n+1}$$

This empirical distribution can be used to estimate the real data distribution (*G*). If

$$\left(y_{(i)},\tilde{G}^{-1}\left(\frac{i}{i+1}\right)\right)$$

$$\tilde{G}^{-1}\left(\frac{1}{n+1}\right)=\hat{\mu}-\frac{\mu}{\xi}\left(1-(-\log\frac{i}{n+1})^{-\xi}\right)$$

yi ≈ GEV assumption is true, then G ≈ G for each. As a result, the graph of the pair of points (G(y(i)),G(y(i))) for i = 1,...,n, will lie along the diagonal line (gradient 1). This is known as the probability graph (probability plot) or P-P Plot. While the chart quantile (Q-Q plot) in the other part is a plot of the pair of points. Similar to the probability graph, if the model assumption is correct, then the points in the QQ plot will be located along the diagonal line.

2.5 The extreme values determination of rainfall and temperature

Extreme values determination can be done in two ways:

1.By taking the maximum values for a certain period, such as weekly or monthly, the observation of these values is considered as extreme values.

2.By taking the values that exceed a threshold value, all values that exceed the threshold are considered as extreme values.

2.6 Return Level

In the practice, the concerned amount/quantity is not only focused on the estimation of the parameters themselves but also on the quantile which is also referred to as return level of the GEV estimators [6,7,14]. Assumed return value of the maximum rainfall and temperature will be used for the validation of the rainfall and temperature data. If F is the distribution of the maximum value for the same period of observation, then;

$$x^k = F^{-1}\left(1-\frac{1}{k}\right)$$

rainfall and temperature will reach a maximum value once (Gilli and Kellezi, 2003) after assumed parameters μ, σ and ξ can be substituted using above equation.

$$x^k = \begin{cases} \hat{\mu} - \dfrac{\hat{\sigma}}{\hat{\xi}}\left(1-\left[-\ln\left(1-\dfrac{1}{p}\right)\right]^{-\xi}\right) & ;\xi \neq 0 \\ \hat{\mu} - \hat{\sigma}\ln\left[-\ln\left(1-\dfrac{1}{p}\right)\right] & ;\xi = 0 \end{cases}$$

III. RESULT

Time series is an observation sequence used to measure the scale of time and space. The data, which is tested periodically, is the maximum rainfall data which consists of monthly rainfall data for 29 years (1984-2012) and temperature data for 16 years (1995-2011) from taken from Pattimura Ambon weather station as in Figure.1 below.

It appears in the Boxplot on annual data, that the average value which is indicated by the line in the middle of the box is always in the down position. The pattern is read that most of the data are on the left side, while others have high extreme value with the advent of the '*'. Characteristics of the data are not symmetrical with the proportion of more data on the left side; it is assumed that the rainfall data will follow a generalized extreme value distribution while the temperature data follows the normal distribution. of the work or suggest applications and extensions.

From the table above, the results of the data distribution test with with error rate (α) of 0.05 will give a result that the tested distribution equivalent with the data distribution, if the calculated probability value is greater than 0.05. Like the model fit test results conducted by the three models above, the acceptable distribution test is the Kolmogorov-Smirnov, the GEV distribution have a probability = 0.781 and has the highest value compared to other distributions. So that the test results have given sufficient proof that the rainfall data follows the GEV distribution with mean (μ) = 218.32, standard deviation (σ) = 170.78 and shape (k) = 0.0718. While temperature follows Normal distribution; probability = 0.455 and has the highest value compared to other distributions. So that the test results have given sufficient proof that the temperature follows a normal distribution with mean = 26.537, standard deviation = 1.201.

Several diagnostic plots are done inferential test to assess the accuracy of the model as shown in Figure 4 and 6. Points on both the Q-Q and P-P plot lay along the diagonal line, an indication that the GEV model is suitable for rainfall and Normal for temperature. If this image can establish a pattern of straight lines, it can be concluded that the distribution matches the distribution of research data. Suitability distribution based on the results of this test will be illustrated through "fit data" towards histograms and cumulative distribution. While the empirical CDF graphs are used to evaluate the fit data from distribution data compared to the sample distribution cumulatively in which both probability values gives a similar shape on both axes.

3.1 Return level

The return level analysis of rainfall and temperature in the city of Ambon gives an idea of how big a maximum value expected to be surpassed on average once in certain period. Return level values of rainfall and temperature are obtained and used as a benchmark for forecasting the occurrence of rainfall and maximum temperature with certain Probabilities. The estimated value of the GEV parameters for the rainfall and temperature results in the return level[11] for 100 years as the table below:

IV CONCLUSION

1.Frequently, from several analyzed extreme data types can meet certain conditions, which are modeled with the GEV distribution, GEV distribution family has three subfamilies, namely Gumbel, Fréchet and Weibull. This distribution is characterized by three parameters: the location, scale, and shape parameter.

2.From the result of distribution analysis and estimation, the histogram of rainfall data in this study has a generalized extreme value distribution pattern while the temperature data in this study has a normal distribution pattern.

3.Suitability distribution based on the test results is described through "*fit data*" to histograms and cumulative distribution. The empirical cdf graphs are used to evaluate the data fit from data distribution which compared from sample distribution cumulatively in which the value of both probabilities gives a similar shape on both axes.

4.Based on the return level on 100 years from the analysis of the maximum temperature, it increased only for 0.66% or the x^k value = 29.367°C, which means that in return period of 100 years, the temperatures in the study area of Ambon city will pass 29.367°C. Therefore, it is certain that the temperature changes in the study area are not too significant and have extreme phenomenon compared to the one with the maximum rainfall.

Acknowledgements

My grateful thanks go to Prof. Dr. Ir. Mohammad Bisri, MS, Dr.Ir. Aniek Masrevaniah, Dipl.HE. and Ir. Agus Suharyanto, M.Eng, Ph.D. for proofreading the manuscript.

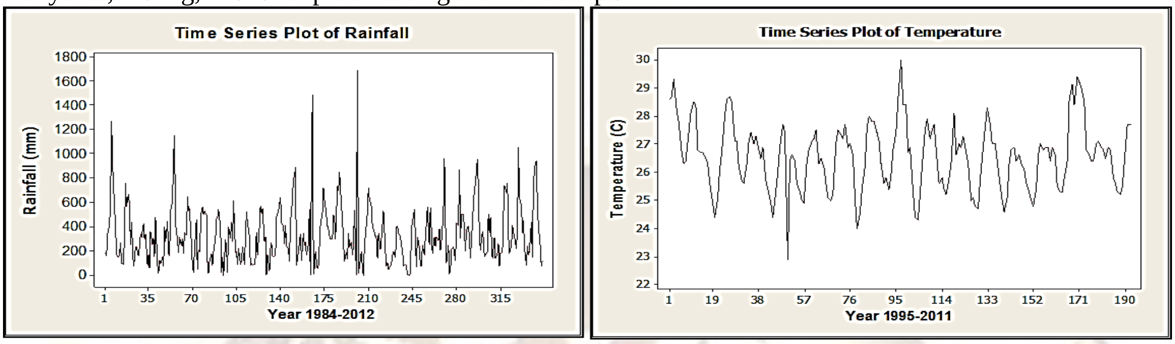

Figure 1 The maximum series data of rainfall and temperature

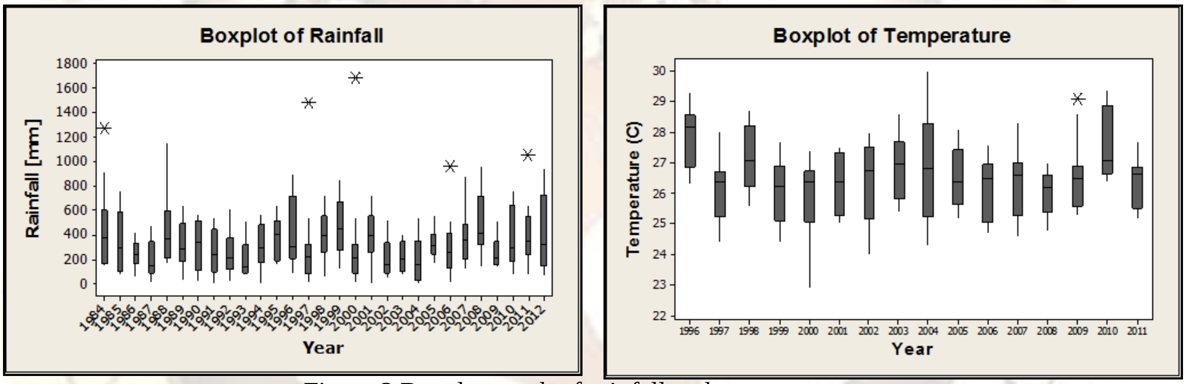

Figure 2 Boxplot graph of rainfall and temperature

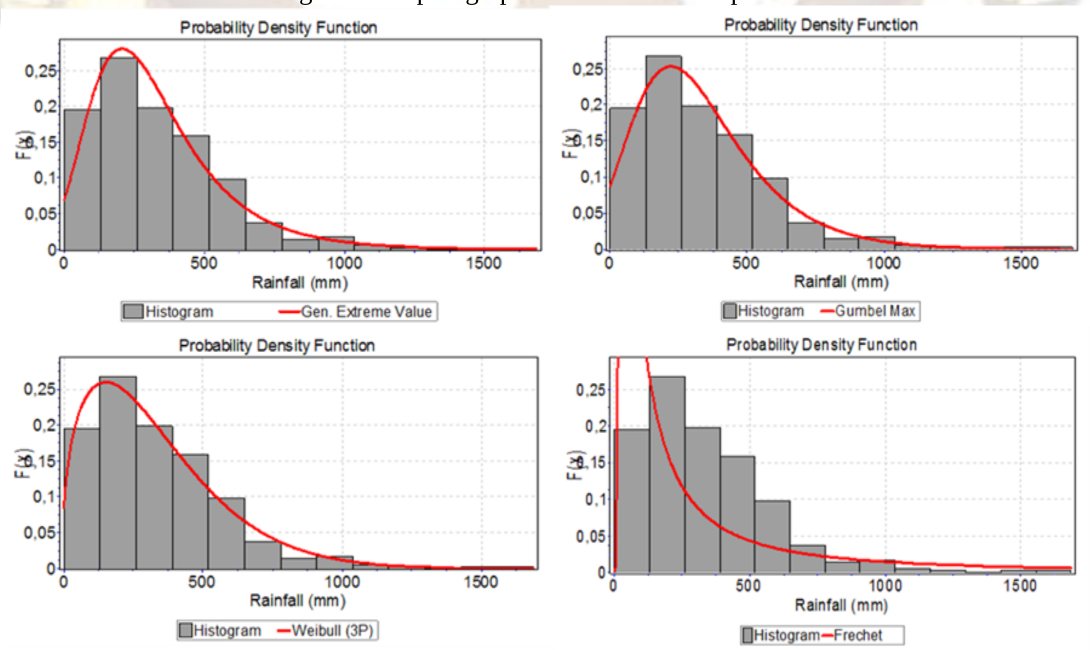

Figure 3 Histogram of rainfall suitability distribution

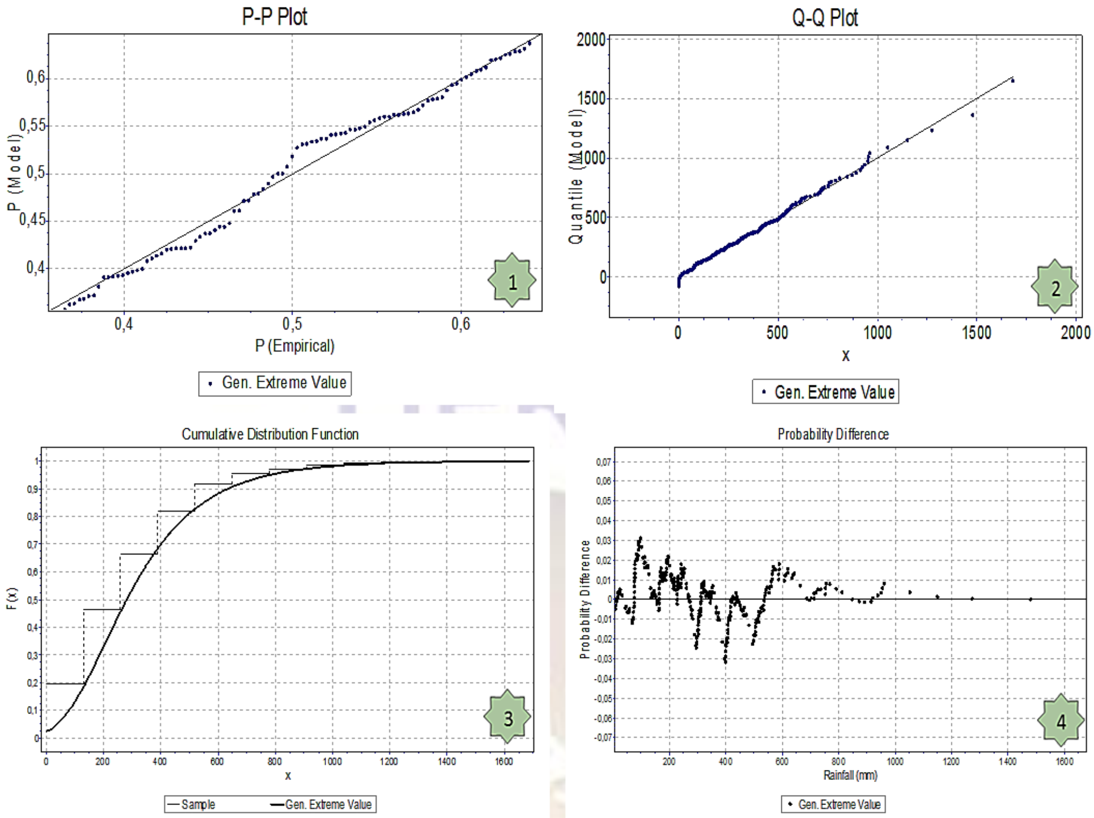

Figure 4 Graph of rainfall diagnostics

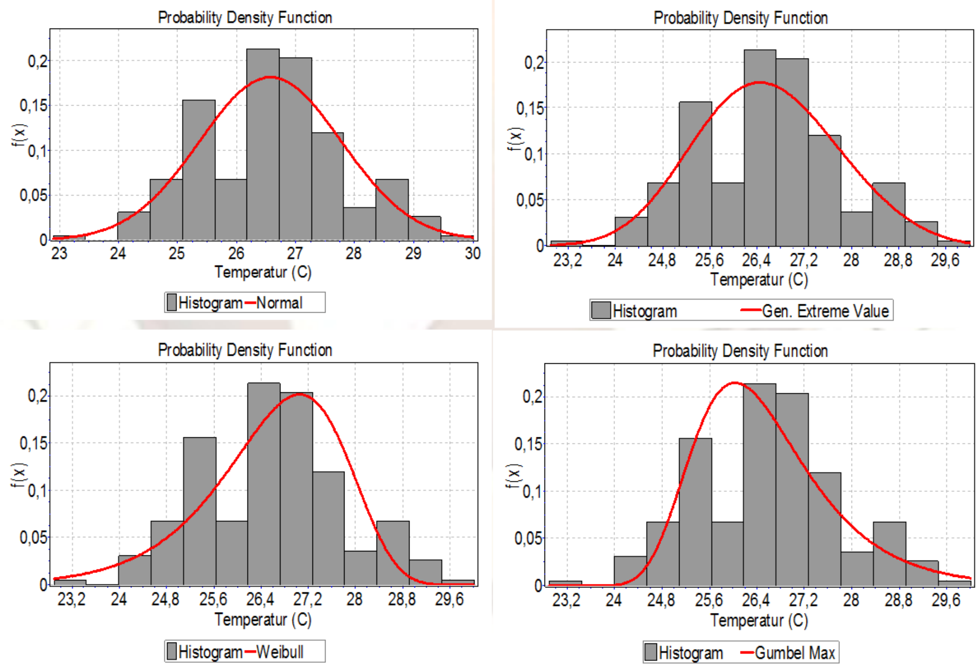

Figure 5 Histogram of temperature suitable distribution

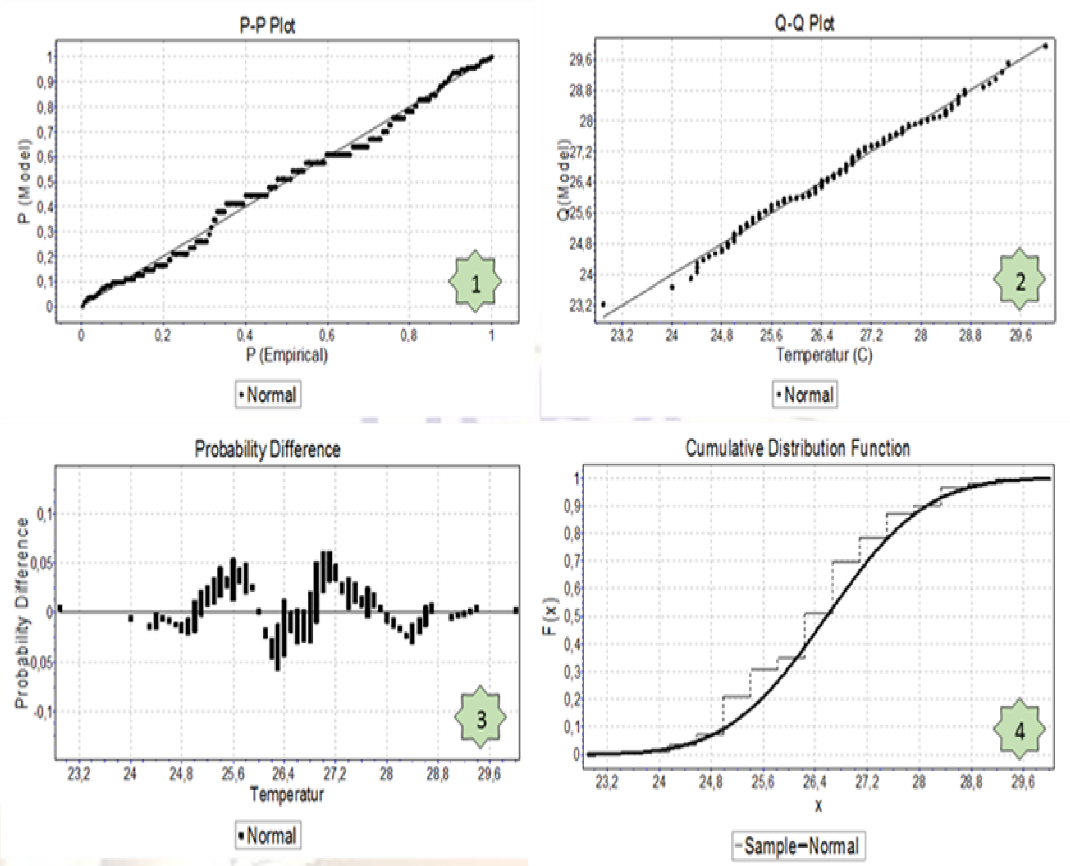

Figure 6 Graph for temperature diagnostics

Table 1 Compatibility Test of rainfall distribution

Distribution	Kolmogorov-Smirnov				Anderson Darling			Chi-Squared			
	Sample	Statistics	Probability	ranking	Sample	Statistics	ranking	(°)Free	Statistics	Probability	ranking
Gama	384	0.0408	0.592	2	384	1.399	2	8	10.054	0.261	2
GEV	348	0.0347	0.781	1	348	0.401	1	8	9.847	0.276	1
Normal	348	0.0981	0.0026	3	348	6.813	3	8	24.309	0.002	3
Weibull	348	0.1487	3.4E-07	4	348	14.999	4	8	95.586	0	4

Table 2 Compatibility Test of temperature distribution

Distribution	Kolmogorov-Smirnov				Anderson Darling			Chi-Squared			
	Sample	Statistics	Probability	ranking	Sample	Statistics	ranking	(°)Free	Statistics	Probability	ranking
Gama	192	0.066	0.347	2	192	0.67	1	7	15.84	0.0265	4
GEV	192	0.073	0.245	3	192	0.687	2	7	15.834	0.0266	3
Normal	192	0.0609	0.455	1	192	0.702	3	7	15.839	0.026	2
Weibull	192	0.098	0.043	4	192	3.807	4	7	12.911	0.074	1

Table 3 MLE distribution estimation for rainfall

Distribution	Parameter				
	α	β	σ	μ	ξ
Gama	1.8617	177.196			
GEV			170.7843	218.3223	0.0718
Normal			241.78	329.9043	
Weibull	0.9111	418.7987			

Table 4 Estimation of temperature distribution for the MLE

Distribution	Parameter				
	α	β	σ	μ	ξ
Gama	489.606	0.0542			
GEV			1.1704	26.13	-0.2437
Normal			1.201	26.537	
Weibull	27.113	27.089			

Table 5 The estimated value of the parameters of GEV and Return level results

Variable	Parameter			Tr(Year)					
	σ	μ	ξ	2	5	10	20	50	100
Max rainfall	170.784	218.322	0.072	280.917	474.489	602.65	725.585	884.712	1003.96
Max temperature	1.1704	26.13	-0.2437	26.5404	27.6004	28.1574	28.6039	29.0769	29.367728

REFERENCES

1. Ramachandra Rao, Shih-Chieh Kao. Statistical Analysis of Rainfall Data Indiana, Joint Transportation Research Program, Purdue University. 2006
2. Ci Yang and David Hill. Modeling stream extremes under non-stationery-time conditions, XIX International Conference on Water Resources CMWR 2012 University of Illinois at Urbana-champaign June 17-22, 2012.
3. Dana Draghicescu, Rosaria Ignaccolo. Modeling threshold exceedance probabilities of spatially correlated time series, Electronic Journal of Statistics Vol. 3 (2009) 149-164 ISSN: 1935-7524 DOI: 10.1214/08-EJS252.2009.
4. Hosking, JRM, and Wallis, JR, and quantile Parameter Estimation for the Generalized Pareto Distribution, Technometrics, 29, 1987, pp. 339-349.
5. Indriati Bisono. Knowing Extreme Data and Distribution, Industrial Engineering Journal, Vol. 13, No.. 2, December 2011, 81-86. 2011
6. James A. Smith and Mary Lynn Baeck, Julia E. Morrison. The Regional Hydrology of Extreme Floods in an Urbanizing Drainage Basin, AMS Online Journal Volume 3, Issue 3 (June 2002).
7. Julia E. Morrison, James A. Smith. Stochastic modeling of flood peaks using the generalized extreme value distribution, Water Resources Research Volume 38, Issue 12, pages 41-1-41-12, December 2002.
8. P. Willems, K. Arnbjerg-Nielsen, J. Olsson, VTV Nguyen. Climate change impact assessment on urban rainfall extremes and urban drainage: Methods and shortcomings, Atmos-02 405; No. of Pages 13.2011
9. Richard W. Katz, Marc B. Parlange, Philippe Naveau. Statistics of extremes in hydrology, Advances in Water Resources 25 (2002) 1287-1304.2002
10. K. Riipunjaii. Shuklla, M.. Triivedii, Manoj Kumar. Proffiiciientt On tthe use off GEV disttriibuttiion, anale. Seria Informatica. Vol. VIII fasc. I - 2010.
11. Slobodan P. Simonovic, Angela Peck. Updated rainfall intensity duration frequency curves for the City of London under the changing climate, Department of Civil and Environmental Engineering The University of Western Ontario London, Ontario, Canada.2010
12. Song Feng, Saralees Nadarajah, Qi Hu. Modeling annual extreme precipitation in china using the generalized extreme value distribution, Journal of the Meteorological Society of Japan, Vol.85.no.5, pp.599-613.2007.
13. Stevan Hadživukovic, RA Fisher and Modern Statistics, International Statistical Institute, 55th Session 2005: Stevan Hadzivukovic, Emilija Nikolic-Doric. , 2005.
14. Xin Liu, Data Analysis in Extreme Value Theory, Department of Statistics and Operations Research University of North Carolina at Chapel Hill. April 30, 2009.