STATISTICAL MODELING OF DAILY EXTREME RAINFALL IN COLOMBO

N.Varathan[1], K.Perera[2] and N.Wikramanayake[3]

[1]Lecturer, Department of Mathematics and Statistics, University of Jaffna, Jaffna, Sri Lanka.
[1]E-mail: varathan10@gmail.com
[1]Telephone: +94-77-9352990

[2]Senior Lecturer, Department of Engineering Mathematics, University of Peradeniya, Peradeniya, Sri Lanka.
[2]E-mail: kanthip@pdn.ac.lk
[2]Telephone: +94-81-2393346

[3]Senior Lecturer, Department of Civil Engineering, Open University, Navala, Sri Lanka.
[3]E-mail: tomwiks@yahoo.com
[3]Telephone: +94-77-7763123

Abstract: The occurrence of heavy rainfalls in Sri Lanka results in significant damage to agriculture, ecology, infrastructure systems, disruption of human activities, injuries and the loss of life. The modelling of extreme rainfall has to be developed to manage the natural resources and the built environment to face the impacts of climate change. The main goal of this study is to find the best fitting distribution to the extreme daily rainfalls measured over the Colombo region for the years 1900-2009 by using the maximum likelihood approach. The study also predicts the extreme rainfalls for return periods and their confidence bands. In this study extreme rainfall events are defined by two different methods based on (1) the annual maximums of the daily rainfalls and (2) the daily rainfalls exceeds some specific threshold value. The Generalized Extreme Value distribution and the Generalized Pareto distribution are fitted to data corresponding to the methods 1 and 2 to describe the extremes of rainfall and to predict its future behaviour. Finally we find the evidence to suggest that the Gumbel distribution provides the most appropriate model for the annual maximums of daily rainfall and the Exponential distribution gives the reasonable model for the daily rainfall data over the threshold value of 100mm for the Colombo location. We derive estimates of 5, 10, 20, 50 and 100 years return levels and its corresponding confidence intervals for extreme daily rainfalls.

Keywords: Annual maximum, Threshold value, Generalized Extreme Value distribution, Pareto distribution, Maximum likelihood estimation

1 Introduction

Extreme rainfall events cause significant damage to agriculture, ecology and infrastructure, disruption of human activities, injuries and loss of lives. In Sri Lanka many areas are affected by the heavy rainfall and the associated floods and land slides. In particular Colombo, the capital of the country faces serious flooding problems in low lying areas due to extreme rainfalls. In order to design measures to reduce the threat of flooding it is necessary to carry out statistical modelling of extreme rainfalls and develop design rainfalls of different return periods.

The statistical analysis of extreme rainfall has been done by the scholars in different locations all over the world. In Sri Lanka, Baheerathan and Shaw(1978) have analyzed Rainfall depth duration frequency studies for Sri Lanka using the annual maximum rainfall depths with 3-,6-,12- and 24-h durations for 19 stations spread over the country. They have analysed data from 8 to 24 years in different stations by fitting Gumbel distribution with maximum likelihood parameter estimation. Dharmasena and Premasiri (1990) studied the same concept but the regionalization technique and linear interpolation of intensities for short durations adopted by Baheerathan and Shaw are not adopted in their study. They used 25 years of data of five regions and considered Gumbel distribution with maximum likelihood estimation technique to fit the data.

In this study, we find the best fitting distribution to extreme daily rainfall by using all available past data from 1900-2009 in Colombo station. We use two techniques to select the sample: one is considering the annual maximums of daily rainfall and the other is selecting exceedances over a specific threshold value. The Generalized Extreme Value distribution (GEV) and the Generalized

Pareto Distribution (GPD) are used to find the best fitting distribution for the above two techniques respectively. The parameters are estimated by maximum likelihood method. Moreover, the outliers are considered and the confidence intervals for predicted extreme rainfalls also developed in our study.

2 Theoretical Framework

2.1 The Extreme Value Distributions

There are three models that are commonly used for extreme value analysis. These are the Gumbel, Frechet, and Weibull distribution functions. The Gumbel is easier to work with since it requires only location and scale parameters, while the Weibull and Frechet require location, scale, and shape parameters. The GEV distribution function is, $H(x) = \exp\{-(1+\xi(x-\mu)/\psi)^{-1/\xi}\}$;where $\psi > 0$ – scale, ξ - shape and μ - location parameter.

According to the value of ξ , H(x) can be divided into following three standard types of distributions:

1. If $\xi \to 0$ (Gumbel Distribution)

   $H(x) = \exp(-e^{-x})$, all x

2. If $\xi > 0$ (Frechet Distribution with $\alpha = 1/\xi$)

   $$H(x) = \begin{cases} 0 & ; x < 0 \\ \exp(-x^{-\alpha}) & ; x > 0 \end{cases}$$

3. If $\xi < 0$ (Weibull Distribution $\alpha = -1/\xi$)

   $$H(x) = \begin{cases} \exp(-|x|^{\alpha}) & ; x < 0 \\ 1 & ; x > 0 \end{cases}$$

2.2 The Exceedances over Threshold

In this technique the data are collected over some specific threshold (cut-off) value. Modelling the extremes under this method enables a more efficient usage of extreme value information than that given by an analysis of annual maxima data, which excludes from inference many extreme events that did not happen to be the largest annual event. As a statistical modeling technique this procedure was popularized by Davison and Smith (1990).

Assuming the daily data to be independent with common distribution function F, the conditional distribution of excesses of a threshold u is determined by, $\Pr(X \le u + y \mid X > u) = 1 - \dfrac{1 - F(u+y)}{1 - F(u)}, y > 0$ .Renormalizing and letting $u \to \infty$ leads to an approximate family of distributions given by, $G(y) = 1 - (1 + \xi(y-u)/\sigma)^{-1/\xi}$ is the Generalized Pareto family. This family describes all non-degenerate limiting distributions of the scaled excess-of-threshold distributions. When $\xi \to 0$ the above generalized Pareto family converges to an Exponential family.

2.3 Return Periods

Briefly, the return period (occurrence interval) can be defined as the average time until the next occurrence of a defined event. When the time to the next occurrence has a geometric distribution, the return period is equal to the inverse of probability of the event occurring in the next time period, that is, $T = 1/P$, where T is the return period, in number of time intervals, and P is the probability of the next event's occurrence in a given time interval.

3 Materials and Methods

The data consists of daily rainfall for the years from 1900 to 2009 for the Colombo location. The data was obtained from the Department of Meteorology, Colombo, which lists the daily rainfalls in millimetres.

We have applied the Univariate Extreme Value Theory to fit the distribution and estimate the return periods for the 110-years (1900-2009) of daily extreme rainfall in Colombo by using the statistical software “GenStat”. The GEV distribution to annual maximums and GPD to rainfall over some specified cutt-off value are considered first, and then by testing the shape parameter the best fitting distribution is identified. Thereafter 95% approximate confidence intervals for return periods are found using the identified model.
By examining the mean residual life plot and the parameter stability plot of sigma, it was decided that a value of 100mm seemed reasonable, as the mean residual life plot was approximately linear for a threshold > 100mm and sigma was stable for values of a threshold > 100mm.

4 Results and Discussion

4.1 Fitting distribution to Annual Maximums of Daily Rainfall

Fitting Generalized Extreme Value Distribution (GEV)

Table 4.1 *Maximum Likelihood Parameter Estimation*

Parameter	Estimate	Standard Error
μ	114.9	5.839
ψ	38.52	4.449
ξ	0.1235	0.1008

After fitting the GEV distribution, we check whether the shape parameter (ξ) is zero or not. (P-value = 0.049<0.05), so the data do not fit the Gumbel distribution. The data fits Frechet distribution (since ξ > 0).

The Table 4.2 gives the return values of the annual maximum rainfall daily and their 95% confidence levels for the return periods 5, 10, 20, 50 and 100 years.

Table 4.2 *Return periods and its 95% Confidence bands*

Probability	Return Period	Return Level	Lower	Upper
0.2000	05	178.4	156.2	200.5
0.1000	10	214.8	181.0	248.6
0.0500	20	253.1	201.4	304.8
0.0200	50	308.0	221.2	394.8
0.0100	100	353.5	229.9	477.0

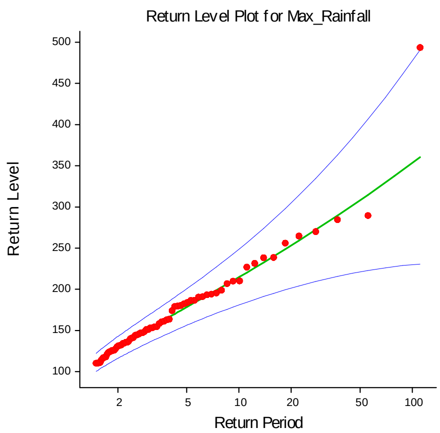

Figure 4.1 *Return Level Plot*

According to the Figure 4.1, it can be seen all the data points are lie within the confidence bands except one point. So we test whether this data point is significant outlier or not, using the Box plot and Grubb's test we found one point was an outlier. This is for the year-1992 annual maximum.

4.2 Fitting GEV Distribution to Outlier Removed Data

After removing the outlier, again we fit the GEV distribution for the annual maximums of daily rainfall data.

Table 4.3 *Maximum Likelihood Parameter Estimation for GEV*

Parameter	Estimate	Standard Error
μ	115.5	5.870
ψ	37.94	4.371
ξ	0.0441	0.1126

For testing the shape parameter ξ = 0, Since the P-value=0.572 > 0.05, there is no evidence to reject the null hypothesis ξ = 0. That is, after removing the outlier annual maximums of daily rainfall data fits Gumbel distribution well.

Table 4.4 *Maximum Likelihood Parameter Estimation for Gumbel*

Parameter	Estimate	Standard Error
μ	116.4	5.496
ψ	38.59	4.165

Table 4.5 *Return periods and its 95% Confidence bands under Gumbel*

Probability	Return Period	Return Level	Lower	Upper
0.2000	05	174.3	155.7	192.9
0.1000	10	203.3	179.3	227.3
0.0500	20	231.0	201.6	260.5
0.0200	50	267.0	230.3	303.6
0.0100	100	293.9	251.8	336.1

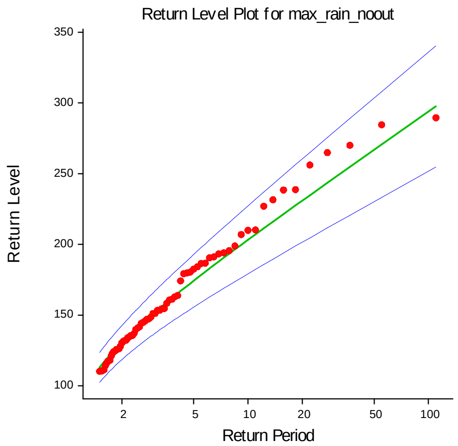

Figure 4.2 *Return Level Plot under Gumbel distribution*

We can observe that the return values of the daily extreme rainfall after removing the outlier is smaller than that of the original data set and the confidence bands width also narrow (Table 4.2 & 4.5). Therefore, it can be said that the Gumbel distribution is the best fit for the annual maximums of daily rain data for the Colombo location.

4.3 Fitting Distribution to Daily Rainfall over a Specified Threshold

Fitting Generalized Pareto Distribution (GPD)

The threshold value of 100mm was found using the Mean Residual life plot and the Stability plot. After removing the outlier, 174 data points were collected using the threshold value of 100mm. By using this collected data, first we fit the Generalized Pareto Distribution (GPD).

Table 4.6 *Maximum Likelihood Parameter Estimation for GPD*

Parameter	Estimate	Standard Error
μ	35.04	5.944
ψ	0.06553	0.1317

The Table 4.6 gives the estimates of the parameters of the GPD distribution using maximum likelihood method. After fitting the GPD distribution, we check the whether the shape parameter (ξ) is zero or not (ie: the data fits the Exponential distribution or not). Since the P-value = 0.3977 >0.05, we don’t have evidence to reject the null hypothesis at 5% level of significance. That is, the data fits the Exponential distribution.

Exponential Distribution:

= 1 - exp(-(y-u)/ 37.47) ; for y > u - the threshold

Threshold u = 100

Based on the identified Exponential distribution we find the return values of the daily rainfall and their 95% confidence levels for the return periods 5, 10, 20, 50 and 100 years. From the Table 4.2.1.2, the 20 year return period is 229.4, which means every 20 year we can expect in average 229.4 mm or more daily extreme rainfall with the probability 0.05.

Table 4.7 *Return periods and its 95% Confidence bands*

Probability	Return Period	Return Level	Lower	Upper
0.2000	05	177.5	160.3	194.7
0.1000	10	203.4	181.0	225.9
0.0500	20	229.4	201.7	257.2
0.0200	50	263.7	228.9	298.6
0.0100	100	289.7	249.5	330.0

Figure 4.3 gives the return periods in years (as the period length was given as 365). Approximate confidence limits for the return periods can be read off the bands.

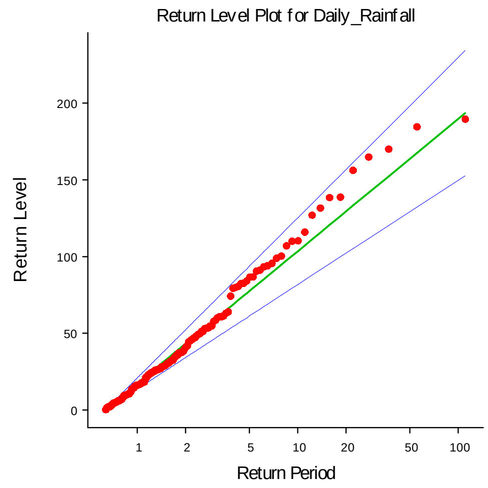

Figure 4.3 *Return Level Plot under Exponential distribution*

Based on the threshold technique the best fitting distribution of the daily rainfall is Exponential with parameter $\psi = 37.47$.

From the Tables 4.5 and 4.7 we can notice, the predicted return values and the confidence levels are very similar in both sampling techniques. When we consider the return period of the outlier 493.7 is nearly 3000 years. So we can’t predict this return value using the above identified results shown in Tables 4.5 and 4.7. Therefore more sophisticated analysis is needed to establish its true return period.

4.4 Goodness of fit Test

In order to test the fitness of the fitted distributions Gumbel and Exponential, the goodness of fit test was carried out. It was observed that empirical distributions agree with the theoretical distributions at the 5% level of significance. Summary of this analysis is given below.

Table 4.8 *Goodness of fit Test Results for Gumbel*

Test	Test Statistic	Critical Value (5%)	Decision
Anderson-Darling	0.237	0.787	Not Significant
Cramer-von Mises	0.033	0.126	Not Significant
Watson	0.031	0.116	Not Significant

Table 4.9 *Goodness of fit Test Results for Exponential*

Test	Test Statistic	Critical Value (5%)	Decision
Anderson-Darling	0.351	0.787	Not Significant
Cramer-von Mises	0.047	0.126	Not Significant
Watson	0.042	0.116	Not Significant

Conclusions

In this study we have performed a statistical modelling of extreme daily rainfall over 110 years in Colombo, Sri Lanka using extreme value distributions under two sampling techniques. Even though the original series of annual maximum daily rainfall data fits the Frechet distribution, the distribution converges to the Gumbel distribution and the predicted values for different return periods and their confidence levels decrease following the removal of the single outlier identified using Grubb's test. Therefore the outlier is more important in this analysis.

We have established the Gumbel and Exponential distributions are suitable models for extreme daily rainfall by considering annual maximums of daily rainfall and daily rainfalls greater than 100mm and checked the adequacy of the models using the goodness of fit test. Finally, we have provided estimates of the return level of daily rainfall and the corresponding 95% confidence intervals for Colombo location in Sri Lanka. These estimates could be used as measures of flood protection.

This paper only provides an initial study of extreme daily rainfall in Colombo. This study can be extended in several ways. One way is to use distributions that are more flexible than the GEV and GPD, such as four parameter Lamda distribution. The other is a more sophisticated analysis of the actual return period of the identified outlier in order to assess its relevance for design.

References

1. Vanniyasingam R. Baheerathan and Elizabeth M.Shaw , Rainfall Depth-Duration-Frequency Studies for Sri Lanka, Journal of Hydrology, 37 , 223-239, 1978.
2. G.T. Dharmasena and S.M. Premasiri, Rainfall Intensity Studies for Sri Lanka, Engineer, December, 1990.
3. Davison A.C and Smith R.L, Models for Exceedances over High Thresholds, Journal of the Royal Statistical Society, 52, No:3, 393-442, 1990.
4. Barbel Finkenstadt and Holger Rootzen, Extreme Values in Finance, Telecommunications, and the Environment, 5th Edition, 2001.
5. Enrique Castillo, Extreme Value Theory in Engineering, Statistical Modeling and Decision Science, Gerald L.Lieberman and Ingram Olkin, editors, Stanford University, Stanford, California.

About the Authors

N.VARATHAN, B.Sc. Hons., Jaffna, M.Sc. Peradeniya, is a Lecturer at the Department of Mathematics and Statistics, University of Jaffna. His research interests are in the areas of Time Series and Regression Analysis.

K.PERERA, B.Sc. Hons., Sri Jeyawardenapura., M.Sc. New York, Albany, Ph.D. New York, Albany is a Senior Lecturer at the Department of Engineering Mathematics, University of Peradeniya. Her research interests are in the areas of Estimation, Regression Analysis and probability Theory.

N.WIKRAMANAYAKE, B.Sc Eng. Hons., Peradeniya, M.Sc. MIT, Ph.D. MIT, MIE (SL), CEng, is a Senior Lecturer at the Department of Civil Engineering, Open University of Sri Lanka. His research interests are in the areas of Hydrology and Coastal Engineering.