JSTEI ONLINE JOURNAL | *Journal of* **SCIENCE, Technology** *&* **Environment Informatics** | **Journal BiNET** Scientific Publication

Rainfall modelling of coastal areas in Bangladesh: extreme-value approach

Nasrin Sultana[1*] and Rezaul Roni[2]

Dept. of Statistics, Bangabandhu Sheikh Mujibur Rahman Agricultural University (BSMRAU), Gazipur-1706, Bangladesh[1]
Dept. of Geography and Environment, Jahangirnagar University (JU), Bangladesh[2]

Article Info.	ABSTRACT
Key words: Generalized extreme value distribution, Pareto distribution and return level Received: 01.12.2015 Accepted: 16.12.2015 Published: 31.12.2015 **Contact author*:** *nasrinisrtdu@yahoo.com*	*A series of rainfall data over 48 years in the period 1966-2014 recorded at six different coastal regions in Bangladesh is modelled using extreme value distributions. In order to reduce destruction and loss of life and property, it is necessary to make proper inference about extreme rainfall. Generalized extreme value distributions (GEV) have been extensively used for this purpose. Fitting annual maximum rainfall according to the block maxima approach. Also generalized pareto distributions (GPD) are fitted to daily rainfall data considering peaks over thresholds (PoT) method. The return levels are the upper tail quantiles that expected to be exceeded once, on average in a given time levels also estimated for different return periods using both models. Assessment of the uncertainty in the estimates of return levels by constructing 95% confidence interval using both delta and profile likelihood methods but due to more accuracy, in this study only present profile likelihood estimate has been discussed.*

I. Introduction

How likely the unusual events are occurring in Bangladesh? How high values of rainfall, wind speeds or sea-level are suspected to risk of flood or tidal surges? The answer is inadequate. Hence, the infrequent or rare events are of particular interest in weather and climate studies. Extreme value theory may provide significant and reliable answer to these questions. It is decisive to have proper inference about extreme events for decision making on rare events to save life and property. The rate of incidence of an

event is extremely large or small with low probability of occurrence is known as extreme events. The stochastic model that characterizes the nature of the tails of probability distributions is extreme value theory. Climatological variables such as wind speed, precipitation, temperature, wave heights, flood as well as credit risk, insurance claims, price fluctuations are studying under climate and weather studies (Cooley, 2005). There are large number of books written by authors as such (Coles, 2001), (Embrechts et al., 1998), (Kotz et al., 2000), (Beirlant et al., 2005); and also recent journals on weather and climatology are good source to learn the theory, scope and application of extreme value approach (Rootzén & Tajvidi, 2006).

Bangladesh is known as a most vulnerable to climate change and natural disaster. Climate of this country is changing every year and become unpredictable. Rainfall which is considered as a significant climate variable and Bangladesh has already experience extreme rainfall over several geographic regions. As consequences of that flood, droughts, cyclones and sea-level rise become very common threats to Bangladesh. Particularly, the southern regions are experienced heavy rainfall during last couple of years that causes major floods. Even, nowadays, in some areas which have not previous record excessive precipitation are experiencing severe flooding and other natural calamities because of these extreme rainfalls. In South Asian region, due to climate change the amount of rainfall increase during the monsoon season while decrease in rainfall in the dry season (Christensen et al., 2007). Bestowing to IPCC (Intergovernmental Panel on Climate Change), glacier melting and more influential monsoons causes around 5% to 6% rise of rainfall over Bangladesh (Parry et al., 2007). As a result more frequent and unadorned floods occur and the southern regions have risk of more coastal flooding and saline interruption. Around 20% of the country distressed due to regular river floods even during extreme year it was increasing up to 68%. According to the Bangladesh National Plan for Disaster Management 2010-2015, there are four types of floods occur all over the country such as overflowing of hilly rivers of eastern and northern Bangladesh cause flash floods, drainage cramming and heavy rains follows rain floods, major rivers usually in the monsoon results monsoon floods and storm surges and tides affected by coastal floods (DMB, 2010).

Serious flooding events already occurred in Bangladesh during in 1988, 1998, 2004 and 2007, 2014, 2015, almost all over the country to varying extent. Also in 1991, 1998, 2000, 2004 and 2007, 2015 there were severe cyclones and tidal surges occurred. These natural hazards are often the tremendous threat to our life and property. Particularly the people in northern Bangladesh are vulnerable by riverbank erosion; homes were flattened or flooded and disrupted power supplies, outbreak of infectious disease and also food insecurity (Daily Star, 2011). According to the International Federation of the Red Cross and Red Crescent Societies river erosion as the largest alarming issue for Bangladesh (New Age, 2011).

Economy of Bangladesh is associated with agricultural activities, around 80% of the population lives in rural areas and is dependent on agriculture by direct or indirect basis. This, erratic rainfall and natural hazards affect the life and economy of Bangladesh. Hence, preparation for future disaster is necessary to avoid circumstantial loss of life and property. Therefore, the keen concern of this paper is to model rainfall data recorded from several metrological stations of Bangladesh and to estimate the return levels for several return periods.

II. Materials and Methods

Statistical model for extreme rainfall

Generalized extreme value distribution: Consider $X_1, X_2, X_3 \ldots \ldots \ldots \ldots . X_n$ is a sequence of independent and identically distributed random variables, having a common continuous distribution function F and M_n represents the maximum value of the process over $n \in N$ time units of observation and can be written as $M_n = \max(X_1, X_2, X_3 \ldots \ldots \ldots \ldots . X_n)$. The Extremal Types Theorem states for some normalizing constant $a_n > 0$ and $b_n \in \mathbf{R}$ the limiting distribution of M_n is shown below.

$$\lim_{n\to\infty} \Pr(\frac{M_n - b_n}{a_n} \le x) = \lim_{n\to\infty} F^n(a_n x + b_n) = G(x)$$

Here, $G(x)$ is non-degenerate distribution function. "The external types" theorem refers that the only possible limits for $G(x)$ are given three types of distributions. The re-parametrization technique by von Mises shows that these three types of distributions can be combined to a one single family of model by changing location and scale. This single distribution popularly known as generalized extreme value (GEV) distribution and can be written as below.

$$G(x; \gamma, \mu, \sigma) = exp\left\{-\left(1 + \gamma\left(\frac{x-\mu}{\sigma}\right)\right)_+^{-\frac{1}{\gamma}}\right\} \dots\dots\dots\dots (2.1)$$

The location (μ) parameter of GEV specifies the center of the distribution, scale (σ) determines the size of deviations of μ and shape (γ) shows how rapidly the upper tail decays. The positive γ implies a heavy tail while negative one implies a bounded tail, and the limit of $\gamma \to 0$ implies an exponential tail. These cases correspond to **type I** (*Fréchet distribution*) and **type II** (*Weibull distribution*) distributions respectively. For $\gamma = 0$ interpreted as the limit which leads to the double exponential or **type III** *(Gumbel).* The end point of distributions is shown below.

$$x > \mu - \frac{\sigma}{\gamma} \text{ for } \gamma > 0 \text{ and } x < \mu - \frac{\sigma}{\gamma} \text{ for } \gamma < 0.$$

The extreme value distribution is used to model the maximum for each block where blocks are partitioned in to non-overlapping periods of equal sizes (say n). Suppose i.i. d. $X_1, X_2, X_3 \dots\dots\dots\dots X_n$ observations are divided in to m blocks of size n. Then, $M_{n,1}, \dots\dots\dots, M_{n,m}$ are the maximum values from each block. The distributions of these maximum observations can be approximated by the asymptotically motivated GEV distribution. In order to obtain good asymptotic approximation, we must not keep block size too small.

Return level: The quantiles or return levels are comfortable way to interpret extreme value instead parameters. For a given value of probability p the estimates of extreme return levels associated with 1/p return period is satisfied the following relation $G(z_p) = 1 - p$. Inverting the GEV distribution, extreme quantiles can be estimated as follows:

$$z_p = \begin{cases} \mu - \frac{\sigma}{\gamma}[1 - (-\log(1 - p))^\gamma], & for\ \gamma \ne 0 \\ \mu - \sigma log[-log(1 - p)], & for\ \gamma = 0 \end{cases} \dots\dots\dots\dots (2.2)$$

Therefore any particular block maima exceeds the value z_p with probability p and the level z_p is expected to be exceeded on average once in every 1/p years.

Profile likelihood confidance interval: The 95% CI is obtained by re parameterization of GEV model considering z_p as a model parameters and maximizing the log likelihood with respect to remaining parameters. The re -parameterization can be done by following way:

$$\mu = z_p + \frac{\sigma}{\gamma}[1 - \{-log(1 - p)\}^{-\gamma}] \dots\dots\dots\dots\dots (2.3)$$

So the GEV model in terms of parameters (z_p, σ, γ) can be obtained by replacement of μ likelihood.

Generalized Pareto distribution: The peaks over threshold technique are usually applied to model daily data. Assuming the daily data $\{X_1, X_2, X_3, \dots\dots\}$ is independent sequence of with common distribution function F. Since this approach is considering only the values that are exceed the high threshold u then the conditional distribution of excesses of a threshold u is determined as shown below.

$$Pr((X > u + x)/X > u) = \frac{1 - F(u+x)}{1 - F(u)}, x > 0 \dots\dots\dots\dots\dots (2.4)$$

The generalized Pareto distribution (GPD) is obtained by using GEV as an approximation and for large value of u the distribution function of $(X - u)$, conditional on $x > u$ approximately follows:

$$H(x) = \begin{cases} 1 - \left(1 + \gamma \frac{x}{\tilde{\sigma}}\right)^{-1/\gamma} & for\ \gamma \neq 0 \\ 1 - exp\left(-\frac{x}{\tilde{\sigma}}\right)^{-1/\gamma} & for\ \gamma = 0 \end{cases} \quad \text{............... (2.5)}$$

Defined on $\left\{x : x > 0\ and\ \left(1 + \gamma \frac{x}{\tilde{\sigma}}\right) > 0\right\}$ where, $\tilde{\sigma} = \sigma + \gamma(u - \mu)$. The family of distribution defined by equation 2.5 known as generalized Pareto distribution family. There are two methods have been used to select appropriate threshold for models namely, threshold range plots and mean residual plots. The return level of GPD can be written as below.

$$x_m = \begin{cases} u + \frac{\tilde{\sigma}}{\gamma}[(m\eta_u)^\gamma - 1] & \gamma \neq 0 \\ u + \tilde{\sigma} \log(m\eta_u) & \gamma = 0 \end{cases} \quad \text{........... (2.6)}$$

Provided for sufficiently large m to confirm that $x_m > u$

Poisson-gpd model for exceedances: The Poisson-GPD model is used to explain the connection between the PoT and Block Maxima methods that can be explain if the number of exceedances over u per year follow Poisson distribution with mean λ, and if the exceedance distribution falls within the generalized Pareto family with shape parameter γ, then the annual maximum distribution falls within the GEV family with the same shape parameter. Another one-one relation between (λ, σ) and $(\mu, \tilde{\sigma})$ that is the location and scale parameters of the annual maximum GEV distribution. The essential benefit of inference on the annual maximum distribution is now derived from a greater amount of information that increased precision.

Extremal index: Smith *et al.* (1997) used geometric distribution to model the cluster length of low minimum daily temperatures. The probability mass function of Geometric distribution is defined as below.

$$f(t, \theta) = \theta(1 - \theta)^{t-1}, \qquad t = 1, 2, \dots$$

Here, θ is known as extremal index and the reciprocal of the parameter θ is mean and used to measure the tendency of clusters of extreme events under stationary process. In this paper, we model the time lag between two extreme rainfalls over threshold u as a geometric distribution to checks stationarity of the rainfall series. More details can be found with Coles *et al.* (2001).

Data: The daily maximum rainfall data recorded at six meteorological coastal stations were collected from the Bangladesh Meteorological Department (BMD). Each station signifies one district (mid-level administrative unit in Bangladesh). There is some missing data the early periods and in 1971(liberation war period), we omit these missing data before analysis. All estimation and calculation in empirical analysis are implemented in R with packages *fitdistrplus*, *in2extRemes*.
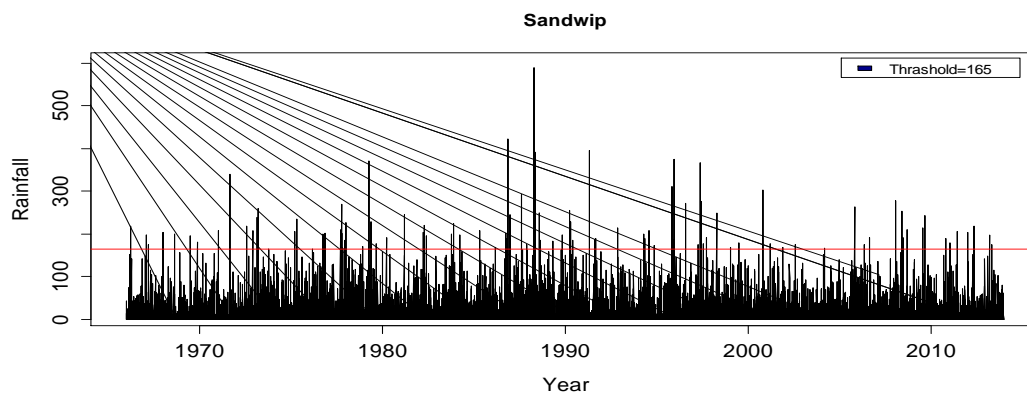
Figure 01. Daily rainfall at Sandwip from 1966-2014

Figure 01 represents the daily rainfall at Sandwip over the period 1966-2014 and we see some extreme rainfall over the periods. Similarly, we have same comment from the other stations rainfall plot which is not presented here for simplicity.

III. Results and Discussion

Daily maximum rainfall data recorded from six different stations over period 1966-2014 are investigated in this study. Trend and stationary are checked before modeling the rainfall data. These can be done by fitting poisson distribution and geometric distribution respectively. The annual maximum rainfall data of each can be modeled as GEV separately in block maxima approach. Another useful model in the extreme value theory is GPD model that used to model daily rainfall data. Here, also GPD is fitted separately for all stations. The peaks over threshold method apply in GPD modeling approach. More clearly the exceedances over a fixed threshold are modeled as GPD. So, we have to be more careful in selecting a threshold. Both mean residual plot and threshold range plots are applied to find appropriate threshold. Different model checking tools like density plots, qq-plots are applied to the fitted model to check suitability of series to the data for both models. The return levels of 10, 20, 50 and 100 years return periods are estimated. These statistics are very useful measure to have approximate idea about the risk of having that level rainfall next 10, 20, 50 or 100 years. Maximum likelihood estimation (MLE) method is used to estimate parameters of GEV distribution are presented in the following Table 01.

For station Barisal, the estimated shape (both GEV and GPD) is negative, i. e., $\gamma < 0$ which refers to **type II**, Weibull classes and the support of the GEV distribution is bounded but the estimated 95% CI (-0.267, 0.209). The strength of evidence from the data for a bounded distribution is not strong as 95% CI extends above zero. For station Bhola, the estimated shape from GEV is negative but close to zero. Moreover, from GPD fit the shape is positive refers with no upper bound. For station Cox's Bazar, the estimated shape (both GEV and GPD) is negative which refers to **type II**, Weibull classes and the support of the GEV distribution is bounded. The estimated 95% CI of shape parameter is (-0.282, -0.025) and the upper end point is $\hat{\mu} - \frac{\hat{\sigma}}{\hat{\gamma}}$ = 565.51. Hence, it can be interpreted as the predicted future rainfall at Cox's Bazar for any return period should never be greater than 565.51 mm. The shape parameters of both GEV and GPD obtained by fitting rainfall at Khulna are approximately equal and positive. The rainfall follows GEV distribution of **type I** (Fréchet family) with no upper bound. The standard errors of estimated parameters of GPD are larger compared to GEV model. The rainfall at Satkhira follows **type II** GEV as the MLE of shape parameters is negative. We can estimate the upper end point $\hat{\mu} - \frac{\hat{\sigma}}{\hat{\gamma}}$ = 1279.09 mm but 95% CI of shape is (-0.222, 0.102) are above zero, so the inference

is not so strong. Furthermore, MLE of GPD shape is positive indicating no upper end point. The rainfall at Sandwip follows **type I** GEV as the MLE of shape parameters is negative and MLE of GPD shape also is positive indicating no upper end point.

Table 01. Summary of GEV and GPD fit

Station	GEV			GPD	
	Location ($\hat{\mu}$)	Scale ($\hat{\sigma}$)	Shape($\hat{\gamma}$)	Scale ($\hat{\sigma}$)	Shape($\hat{\gamma}$)
Barisal	119.21(6.43)	39.23(4.70)	-0.030(0.121)	39.51(5.19)	-0.0396(0.096)
Bhola	123.27(8.75)	57.09(5.82)	-0.006(0.053)	38.66(6.34)	0.19(0.122)
Cox's Bazar	177.25 (9.23)	9.54(6.049)	-0.153(0.066)	44.53(6.05)	-0.013(0.091)
Khulna	100.64(6.31)	38.38(5.12)	0.250(0.117)	24.59(3.597)	0.327(0.117)
Satkhira	97.35(7.09)	44.32(4.90)	-0.060(0.083)	37.38(6.88)	0.029(0.133)
Sandwip	178.42(7.32)	43.19(6.04)	0.253(0.138)	54.50(8.08)	0.096(0.107)

The model diagnostic plot looks reasonable for rainfall at Sandwip (figure 02). Similar way we made diagnostic plot for other stations that also looks good, are not presented in this study.
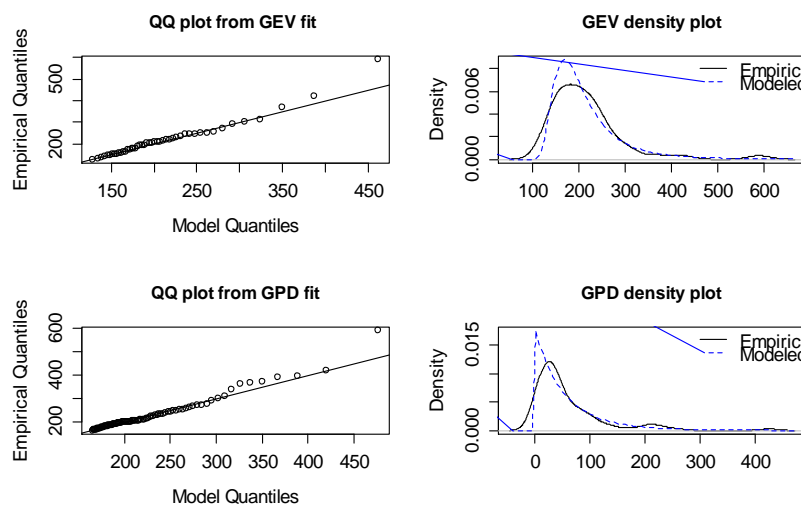
Figure 02. Model diagnostic QQ plots and density plots at Sandwip
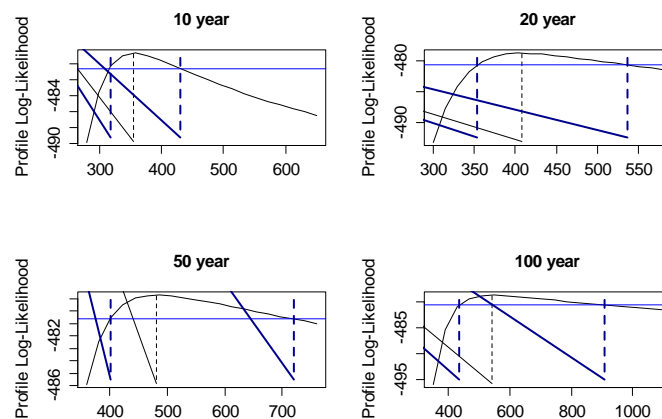
Return level of rainfall

The maximum likelihood estimate of return levels estimated for 10-year, 20-years, 50-year and 100-year return periods and 95% confidence interval (inside bracket) for every station are shown in Table 02. The values of the 95% confidence interval estimated from profile likelihood. The estimated return levels from GPD and GEV are approximately equal for that we are presenting the return levels from GEV. The maximum 100 years return levels of rainfall is approximately 554.21 mm at Sandwip. So, the expected rainfall 554.21 mm is to be exceeded on average once every 100 years with probability 1/100. The minimum return levels compared to other stations is 193 mm for 10 years return period is at Satkhira. The expected return levels for different return periods are maximum at Sandwip.

Table 02. Estimated return levels with confidence interval

Stations	10 year	20 year	50 year	100 year
Barisal	205.19 (194.7368,239.963)	228.95 (205.11, 283.83)	259.37 (231.58,350.60)	281.67 (239.91,408.91)
Bhola	248.79 (231.58, 318.49)	297.43 (248.44, 425.73)	372.27 (291.54, 639.26)	438.10 (342.11, 875.30)
Cox's Bazar	286.47 (263.70, 326.38)	316.04 (290.53,378.97)	354.72 (314.67,461.07)	383.69 (342.11,534.91)
Khulna	222.44 (186.34, 303.77)	277.81 (219.47, 432.87)	373.19 (267.79,710.43)	466.92 (310.92,1051.44)
Satkhira	193.04 (173.68,237.56)	222.17 (192.81, 298.89)	262.72 (220.53, 408.26)	295.02 (252.63, 519.23)
Sandwip	309.35 (318.95, 431.29)	369.58 (353.69, 536.16)	465.75 (402.11, 720.43)	554.21 (436.84, 906.72)

Profile likelihood CI usually provides better accuracy in estimating confidence interval, which are slightly different from delta method. So, here we only present profile likelihood CI.

The horizontal line of figure 03 is at maximum profile likelihood value minus the 0.975 quantile of χ_1^2 distribution function. Therefore, all the return levels associated profile likelihood above the lower horizontal line in the figure are between 95% Confidence region and the values below the line are outside of 95% region. Approximate CI is the points that cross the horizontal line and the blue vertical line is the estimated CL and middle black line is MLE. Similar way we can plot the return levels of other stations.

Figure 03. Profile likelihood CI of 100 years return levels at Sandwip

IV. Conclusion

In modeling the statistical behavior of extreme rainfall recorded from six different stations over 48 years, two approaches are used under this study. In block maxima approach, the generalized extreme value distribution (GEV) is used to model annual maxima rainfall data. For peaks over threshold (PoT) method, generalized pareto distribution (GPD) is applied to model the daily rainfall data. These two distributions are interrelated through Poisson-GPD model and provide similar inference about extreme values. The return level is obtained to predict the rainfall for the long run in the future profile likelihood CI also estimated. The plots of profile likelihood are usually make the certainty about the correctness of the resulting CI. The asymmetry shape of long return period remarks its correctness. Though the plots

for a particular fit is tough but provide better accuracy particularly long return period. Throughout this study, how extreme value theory used as a significant tool in describing extreme events is explained. In this paper, two models are considered for daily rainfall data as well as annual maximum rainfall data. The both model are well fitted and interrelated. Station Sandwip has maximum return levels compared to other stations. Also, Bhola and Khulna have large return levels. Thus, these stations have risk of having extreme rainfall in the next 10, 20, 50 and also 100 years return periods and having probability of natural disaster due to these extreme rainfalls. Also, through the extreme value model at station Cox's Bazar the upper bound of rainfall is found. This study would play as advantageous in understanding extreme rainfall in coastal rainfall areas in Bangladesh. This lends strong support to take proper caution against the natural disasters due to extreme rainfall and save the life and property of coastal areas.

Acknowledgement

V. References

[1]. Beirlant, J., Dierckx, G. & Guillou, A. (2005). Estimation of the extreme-value index and generalized quantile plots. *Bernoulli*, 11(6), 949-970. http://dx.doi.org/10.3150/bj/1137421635

[2]. Christensen, J. H., Hewitson, B., Busuioc, A., Chen, A., Gao, X., Held, R. & Dethloff, K. (2007). Regional climate projections. Climate Change, 2007: The physical science basis. Contribution of working group to the fourth assessment report of the intergovernmental panel on climate change. University Press, Cambridge, Chapter 11, 847-940.

[3]. Coles, S., Bawa, J., Trenner, L. & Dorazio, P. (2001). An introduction to statistical modeling of extreme values (Vol. 208). London: Springer. http://dx.doi.org/10.1007/978-1-4471-3675-0

[4]. Cooley, D. S. (2005). Statistical analysis of extremes motivated by weather and climate studies: applied and theoretical advances. Doctoral dissertation, University of Colorado.

[5]. Daily Star (2011). Climate change: Threat to international peace and security. Available at http://www.thedailystar.net/newDesign/newsdetails.php?nid=198040. Accessed on August 11, 2011.

[6]. Embrechts, P., Kluppelberg, C. & Mikosch, T. (1999). Modelling extremal events. *British actuarial journal*, 5(2), 465-465.

[7]. Kotz, S. & Nadarajah, S. (2000). Extreme value distributions (Vol. 31). Imperial College Press, London.

[8]. DMB (2010). National Plan for Disaster Management (2010-2015), Disaster Management Bureau (DMB), Disaster Management and Relief Division (DM&RD), Dhaka, Bangladesh.

[9]. New Age (2011). River bank erosion, displacement and consequences. Available at http://newagebd.com/newspaper1/op-ed/33554.html. Accessed September 19, 2011.

[10]. Parry, M. L. (Ed.). (2007). Climate Change 2007: impacts, adaptation and vulnerability: contribution of Working Group II to the fourth assessment report of the Intergovernmental Panel on Climate Change (Vol. 4). Cambridge University Press.

[11]. Rootzén, H. & Tajvidi, N. (2006). Multivariate generalized Pareto distributions. *Bernoulli*, 917-930. http://dx.doi.org/10.3150/bj/1161614952

[12]. Smith, R. L., Tawn, J. A. & Coles, S. G. (1997). Markov chain models for threshold exceedances. *Biometrika*, 84(2), 249-268. http://dx.doi.org/10.1093/biomet/84.2.249

CITATIONS

Citation for this article (APA Style):
Sultana, N. & Roni, R. (2015). Rainfall modelling of coastal areas in Bangladesh: extreme-value approach. *Journal of Science, Technology & Environment Informatics,* 02(02), 42–50.

MLA (Modern Language Association)
Sultana, N. & Roni, R. "Rainfall modelling of coastal areas in Bangladesh: extreme-value approach." *Journal of Science, Technology & Environment Informatics,* 02.02 (2015): 42-50.

Chicago/Turabian
Sultana, N. & Roni, R. "Rainfall modelling of coastal areas in Bangladesh: extreme-value approach." *Journal of Science, Technology & Environment Informatics,* 02, no. 02 (2015): 42-50.

Submit article and publish with Journal BiNET.

Key Features:

✓ Faster processing and review of article

✓ Quick review and editorial tasks

✓ International editorial and review board

✓ 21 business day's rapid publication

✓ View and read articles powered by Scribd

✓ Greater audience readership is ensured for all article

✓ Indexing and bibliographic integration of article

✓ Social sharing enabled article

Submit article (or email to submit@journalbinet.com):

http://www.journalbinet.com/article-submission-form.html