# MODELLING EXTREME RAINFALLS USING GENERALIZED ADDITIVE MODELS FOR LOCATION, SCALE AND SHAPE PARAMETERS

SEZER, A. – KAN KILINC, B.* – YAZICI, B.

*Department of Statistics, Science Faculty, Anadolu University*
*(phone: (0222) 3350580, Eskişehir,26470 Turkey*

**Corresponding author:*
*email: bkan@anadolu.edu.tr*

(Received 3rd Jul 2016; accepted 7th Sep 2016)

**Abstract.** This study aims to model the nonlinear relationship between the daily amount of extreme rainfall and significant predictor variables by the Generalized additive models for location, scale and shape parameters (GAMLSS). Statistical modelling of extreme rainfall is an essential means of assessing hydrological impacts of changing rainfall patterns resulting from climate variability. Extreme value theory states that only three types of distributions are needed to model the extreme events (Gumbel, Frećhet and Weibull) for large samples. However we identify the model that best characterizes the behaviour of the extreme rainfall data is the lognormal model with respect to Akaike Information Criteria (AIC). In the simulation study, we propose to approximate the location parameter for the Gumbel (maximum) and Lognormal distributions using cubic splines. Results reveal that the approximated mean function by the GAMLSS modelling converges to the true mean function. Moreover, the bias is decreasing rapidly for the true fixed parameter. Although GAMLSS procedure utilizes extreme rainfall data, the same methodology can be applied to other variables in many areas.
**Keywords:** *generalized extreme value distribution, nonparametric regression, extreme, rainfall, smooth splines*

## Introduction

Fisher and Tippett (1928) introduced the asymptotic theory of extreme value distributions. Gnedenko (1943) provided that under certain conditions, three families of distributions (Gumbel, Frećhet, and Weibull) arise as limiting distributions of extreme values in random samples. Coles (2001) defined a general introduction to Extreme Value Theory (EVT). In the extreme value context, Davison and Ramesh (2000), Chavez-Demoulin and Davison (2005) and Yee and Stephenson (2007) have demonstrated the usefulness of the nonparametric regression.

Extreme value distributions are widely used in risk management, finance, economics, hydrology and many other industries dealing with extreme events. Changes in extreme climate events are particularly thought important due to their impacts on human life. Hosking and Wallis (1997) examined the changes to the frequency and intensity of extreme rainfall events by peak-over-threshold analysis. Kharin and Zwiers (2000) highlighted possible future changes in extremes of daily temperature and their effects on extreme precipitation event. Katz et al. (2002) used EVT in water resource engineering and management studies to obtain probability distribution to fit minima or maxima of the data in random samples. Coles et al. (2003) and Sang and Gelfand (2009) studied extreme value analysis in environmental science.

Floods are one of the most costly types of natural disasters in economic and human terms in all around the world. There is no question that extreme rainfalls have


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1404_635644

tremendous effect on human activities, agricultural activities and water resources. Rainfall patterns have also recognized effect on erosion and water quality. Rainfall extremes were spatially modelled by Lehman et al. (2016) but there are not enough number of studies that focus on extreme rainfall modeling.

In this study, the behavior of the extreme rainfall data is studied and the lognormal model gave the best results with respect to Akaike Information Criteria (AIC). In the simulation study, we propose to approximate the location parameter for the Gumbel (maximum) and Lognormal distributions using cubic splines. Results reveal that the approximated mean function by the GAMLSS modelling converges to the true mean function.

## Methods and Methodology

### *The extreme value theory and Generalized Extreme Value (GEV) distribution*

The extreme value distribution is defined from the limit theorem of Fisher and Tippett (1928) on maxima in a sample data. The class of GEV distributions is very flexible with the shape parameter ($\zeta^{-1}$). The Generalized Extreme Value distribution is given by:

$$G(x;\mu,\sigma,\zeta) = \exp\left\{-[1+\zeta(\frac{x-\mu}{\sigma})]_{+}^{-1/\zeta}\right\} \quad \text{(Eq.1)}$$

Eq. (1) defined on $\{x:1+\zeta(x-\mu)/\sigma>0\}$. Here we let $x_{+}=\max(x,0)$ and $\mu,\sigma,\zeta$ are the location, scale, and shape parameters, respectively. The cases; $\zeta>0$, $\zeta<0$ and $\zeta=0$ correspond to the Frećhet (heavy-tailed), Weibull (short-tailed), and Gumbel (light-tailed) distributions, respectively (Yee and Stephenson, 2007). The distributions associated with $\zeta>0$ include well known fat tailed distributions such as the Pareto, Cauchy, Student-t distributions. If $\zeta=0$ the GEV distribution is the Gumbel class and includes the normal, exponential, gamma and lognormal distributions where only the lognormal distribution has a moderately heavy tail. Finally, in the case where $\zeta<0$, the distribution class is Weibull (Markose and Alentorn, 2005).

### *Generalized additive models*

Friedman and Stuetzle (1981) represented regression surface as a sum of general smooth functions of linear combinations of the predictor variables iteratively in the projection pursuit regression. The Generalized Linear Models (GLM) with a linear predictor involving a sum of smooth functions of covariates is known Generalized Additive Models (GAM) and introduced by Hastie and Tibshirani (1987). Diverse nonparametric regression models and their inference procedures were presented by Ruppert et al. (2003) and Yatchew (2003).

Let $y$ be a response variable, and $x=(x_1, \ldots, x_k)$ be a set of $k$ independent variables, GAM assumes that the mean of the response variable depends on additive independent variables through a nonlinear function. The additive model generalizes the linear model by modeling the expected value of $y$ as


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1404_635644

$$\mathrm{E}(y \mid x) = s(x_1, ..., x_\mathrm{k}) = s_0 + s_1(x_1) + ... + s_\mathrm{k}(x_\mathrm{k}) \quad \text{(Eq.2)}$$

where $s_\mathrm{i}(x)$ are arbitrary univariate smooth functions, $i$=1,…,$k$. Locally polynomial splines, kernel smoothers, and cubic splines are the most extensively studied in GAM. Unlike the general linear model, the additive component $\eta$ is not restricted to be linear but is the sum of smoothing functions as given in Eq. (3)

$$\eta = s_0 + s_1(x_1) + ... + s_\mathrm{k}(x_\mathrm{k}) \quad \text{(Eq.3)}$$

GLM generalizes linear model via a link function g(.) and by allowing the magnitude of the variance of each measurement to be a function of its predicted value.

$$g(\mathrm{E}(y \mid x)) = \eta \quad \text{(Eq.4)}$$

*Generalized additive models for location, scale, and shape (GAMLSS)*

Rigby and Stasinopoulos (2001) introduced generalized additive models for location, scale, and shape (GAMLSS) to overcome some of the limitations with GAM.

GAMLSS assumes independent observations $y_\mathrm{i}$, $i$=1,…, $n$ with the probability density function $\mathrm{f}(y_i \mid \boldsymbol{\theta^i})$ where $\boldsymbol{\theta^i} = (\theta_{i1}, \theta_{\mathrm{i}2}, \theta_{\mathrm{i}3}) = (\mu_i, \sigma_i, \varsigma_i)$ is a vector of three distribution parameters, each of which can be a function of the explanatory variables. The distribution parameters are referred as $(\mu_i, \sigma_i, \varsigma_i)$. The first two distribution parameters are defined as location and scale parameters while the last distribution parameter is characterized as shape parameter. In GAMLSS the exponential family distribution assumption is relaxed and replaced by the general distribution family including highly skew distributions (Rigby and Stasinopoulos, 2001; Rigby and Stasinopoulos, 2005). Let $\mathrm{y}' = (y_1, y_2, ..., y_\mathrm{n})$ be the $n$ length of vector of response variable.

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \mathbf{X}_k \beta_k + \sum_{j=1}^{J_k} \mathbf{Z}_{j_k} \gamma_{j_k} \quad \text{(Eq.5)}$$

For $k$=1,2,3, let $g_k(.)$ be known monotonic link functions relating to the distribution parameters to explanatory variables by

$$g_1(\boldsymbol{\mu}) = \boldsymbol{\eta}_1 = \mathbf{X}_1 \beta_1 + \sum_{j=1}^{J_1} \mathbf{Z}_{j_1} \gamma_{j_1}$$

$$g_2(\boldsymbol{\sigma}) = \boldsymbol{\eta}_2 = \mathbf{X}_2 \beta_2 + \sum_{j=1}^{J_2} \mathbf{Z}_{j_2} \gamma_{j_2}$$

$$g_3(\boldsymbol{\varsigma}) = \boldsymbol{\eta}_3 = \mathbf{X}_3 \beta_3 + \sum_{j=1}^{J_3} \mathbf{Z}_{j_3} \gamma_{j_3} \; .$$


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1404_635644

where $\boldsymbol{\mu,\sigma,\zeta}$ and $\boldsymbol{\eta}_k$ are vectors of length *n*, $\beta_k' = (\beta_{1k}, \beta_{2k}, ..., \beta_{J_k'k})$ is a parameter of vector of length $J_k'$, $\mathbf{X}_k$ is a fixed known design matrix of order $n \times J_k'$, $\mathbf{Z}_{jk}$ is a fixed known $n \times q_{jk}$ design matrix and $\gamma_{jk}$ is a $q_{jk}$ dimensional random variable which is assumed to be distributed as $N_{q_{jk}}(\mathbf{0}, \mathbf{G}_{jk}^{-1})$, where $\mathbf{G}_{jk}^{-1}$ is the generalized inverse of a $q_{jk} \times q_{jk}$ symmetric matrix $\mathbf{G}_{jk} = \mathbf{G}_{jk}(\boldsymbol{\lambda}_{jk})$ which may be depend on a vector of hyperparameters $\boldsymbol{\lambda}_{jk}$.

The model in Eq. (5) can be formulated in a semi-parametric way. Let $\mathbf{Z}_{jk} = \mathbf{I}_n$ where $\mathbf{I}_n$ is an $n \times n$ identity matrix, and $\gamma_{jk} = \mathbf{s}_{jk} = s_{jk}(x_{jk})$ for all combinations of *j* and *k* in Eq. (5), then the semi-parametric additive model is given by $g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \mathbf{X}_k \beta_k + \sum_{j=1}^{J_k} s_{jk}(x_{jk})$ where to abbreviate the notation use $\boldsymbol{\theta}_k$ for k=1,2,3 to represent the distribution parameter vectors $\boldsymbol{\mu,\sigma,\zeta}$ and where $x_{jk}$ for *j*=1,2,…,$J_k$ are also vectors of length *n*. The function $s_{jk}$ is an unknown function of explanatory variable $X_{jk}$ and $\mathbf{s}_{jk} = s_{jk}(x_{jk})$ is the vector which evaluates the function $s_{jk}$ at $x_{jk}$. The parametric vectors $\beta_k$ and the additive terms $s_{jk}$ are estimated by maximizing a penalized likelihood function $l_p$ (Rigby and Stasinopoulos, 2001). A Newton-Raphson algorithm was used to maximize the penalized likelihood function. The additive terms in the model are fitted by using a backfitting algorithm.

Model selection can be performed using Global Deviance (GD) or Akaike Information Criteria (AIC); $GD = -2\ell(\hat{\boldsymbol{\theta}})$, $AIC = -2\ell(\hat{\boldsymbol{\theta}}) + 2q$, and *q* is the total degrees of freedom used in all linear parametric and nonparametric terms in the model. The log-likelihood function is given by

$$\ell = \sum_{i=1}^{n} \log f(y_i \mid \boldsymbol{\theta}^i) \qquad \text{(Eq.6)}$$

For each fitted GAMLSS model the (randomized) quantile residuals can be used to check to adequacy of the model. The randomized quantile residuals are given by

$$\hat{r}_i = \Phi^{-1}(\hat{u}_i) \qquad \text{(Eq.7)}$$

where $\hat{u}_i = F(y_i \mid \hat{\boldsymbol{\theta}}^i)$ if $y_i$ is continuous. Otherwise, it is a random value from a uniform distribution in the interval $[F((y_i - 1) \mid \boldsymbol{\theta}^i), F(y_i \mid \boldsymbol{\theta}^i)]$ and $\Phi^{-1}$ is the inverse cumulative distribution function of a standard normal variable.

## Simulation study and results

In this study, the location parameter is assumed to be smooth on an interval [a, b] in the **X** domain and the scale parameter is constant for both Gumbel and Lognormal distribution.


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1404_635644

Markov Chain Monte Carlo (MCMC) simulation is conducted to examine the finite sample performance of the GAMLSS method. The response variable is considered as Gumbel and Lognormal distributed since they have been commonly used for quantifying risk associated with extreme rainfall data.

The mean integrated squared error (MISE) is calculated as the performance of the estimated functions and given

$$\text{MISE}=\frac{1}{\text{N}}\sum_{\text{i=1}}^{\text{N}}(\hat{\text{f}}(\text{x}_\text{i})-\text{f}(\text{x}_\text{i}))^2 \qquad \text{(Eq.8)}$$

The simulation algorithm runs in the following order:

1. A sample $x_1,...,x_n$ is generated from a uniform distribution on [0,1] with sample size $n$ (n=20,50,75,100).
2. The bump function (A.1), true function, is used to generate the location parameter $\mu(x_i)$, i=1,…,$n$ for Gumbel distribution.
3. The cyclic function (A.2), true function, is used to generate the location parameter $\mu(x_i)$, i=1,…,$n$ for Lognormal distribution.
4. The scale parameter is set to 1 and 4 both for Gumbel and Lognormal distributions (A.3).
5. The response values, $y_i$ are generated from both Gumbel distribution and Lognormal distribution with the parameters $\mu(x_i)$ for i=1,…,$n$, $s=1$ and $s=4$
6. Cubic splines are used to approximate the mean function of the location parameter and it is donated by $\hat{f}$.
7. GAMLSSs are fitted to each sample ($n$ =20,50,75,100) for 1000 repetitions.
8. After the fit, MISE in Eq. (8) and $\hat{\sigma}$ are evaluated.

Simulation study reveals that MISE decreases as the sample size increases for both the Gumbel and lognormal models. MISE is relatively large when the sample size is small ($n$=20) but decreases significantly when the sample size increases to 100. Furthermore, estimated functions of the location parameter converge to the true function with a very little bias when the sample size is 100.

The results for the maximum likelihood estimation of the scale parameter ($\sigma$) and MISE scores were summarized in *Table 1-4.*

***Table 1.*** *Gumbel Distribution with true $\sigma$ =1 (1000 repetitions)*

| ***n*** | **MISE** | $\hat{\sigma}$ | **AIC** | **Deviance** |
|---|---|---|---|---|
| 20 | 0.268 | 0.805 | 66.59 | 54.59 |
| 50 | 0.094 | 0.921 | 161.88 | 149.88 |
| 75 | 0.060 | 0.946 | 240.80 | 228.79 |
| 100 | 0.043 | 0.964 | 320.66 | 308.66 |


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1404_635644

***Table 2.*** *Gumbel Distribution with true $\sigma$ =4 (1000 repetitions)*

| *n* | MISE | $\hat{\sigma}$ | AIC | Deviance |
|---|---|---|---|---|
| 20 | 4.096 | 3.226 | 122.20 | 110.19 |
| 50 | 1.486 | 3.699 | 300.83 | 288.83 |
| 75 | 0.971 | 3.778 | 448.52 | 436.52 |
| 100 | 0.682 | 3.849 | 597.41 | 585.41 |

***Table 3***. *Lognormal Distribution with true $\sigma$ =1 (1000 repetitions)*

| *n* | MISE | $\hat{\sigma}$ | AIC | Deviance |
|---|---|---|---|---|
| 20 | 0.212 | 0.833 | 101.270 | 89.277 |
| 50 | 0.082 | 0.931 | 246.194 | 234.202 |
| 75 | 0.054 | 0.956 | 366.786 | 354.794 |
| 100 | 0.040 | 0.966 | 489.485 | 477.493 |

***Table 4.*** *Lognormal Distribution with true $\sigma$ =4 (1000 repetitions)*

| *n* | MISE | $\hat{\sigma}$ | AIC | Deviance |
|---|---|---|---|---|
| 20 | 3.240 | 3.330 | 156.261 | 144.260 |
| 50 | 1.331 | 3.715 | 385.855 | 373.839 |
| 75 | 0.889 | 3.812 | 573.270 | 561.270 |
| 100 | 0.644 | 3.879 | 765.970 | 754.000 |

As expected, the estimated scale parameter $\hat{\sigma}$ approaches to the fixed true values ($\sigma$ =1 and $\sigma$ =4) as the sample size increases. Overall the results indicate that cubic spline approximation of the mean of the location parameter by GAMLSS performs quite well. Indeed, GAMLSS provides efficient and consistent maximum likelihood estimator for the scale parameter.

## *Analysis of extreme rainfall data*

Daily extreme rainfall (Y) was examined for each month between 1991-2010. Data were obtained from Turkish Institute of Meteorology for Trabzon, located in the black sea region in the northern part of Turkey. It is considered that humidity ($X_1$), air pressure ($X_2$), wind speed ($X_3$), and temperature ($X_4$) may have significant effect on extreme rainfall. The distribution of the extreme rainfall was plotted in *Figure 1* and it has a long right tail. Since extreme rainfall data are positively skew, the standard multivariate regression modeling is not appropriate. Indeed, the value of $R^2_{adj}$=0.047 suggests that the rainfall data should be modeled in a nonlinear form with the predictor variables.


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1404_635644

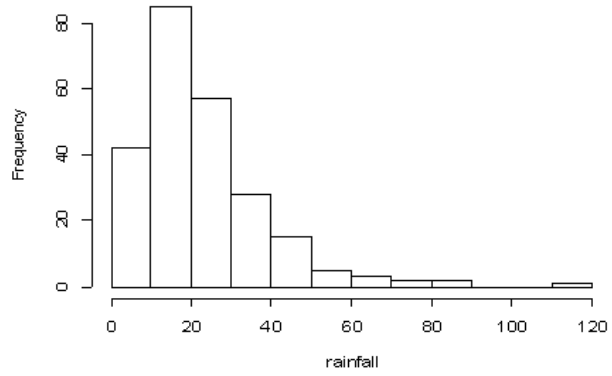


***Figure 1.*** *Histogram of daily extreme rainfall data from 1991 to 2010*

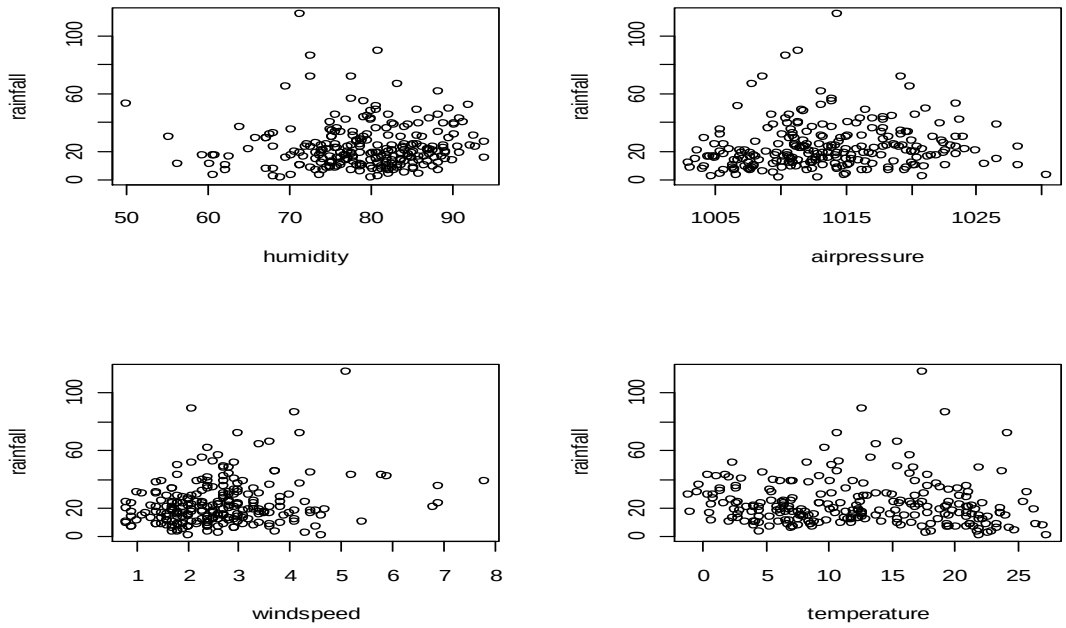


***Figure 2.*** *Scatter plots of rainfall versus predictor variables*

Since the histogram of extreme rainfall data follows a positive skewed distribution, Gumbel and Lognormal distributions were fitted to the rainfall data. In *Figure 2* the scatter plots of rainfall versus predictor variables are presented. AIC is used to obtain the best model among the all possible others. The proposed model for the Gumbel distribution is given in Eq. (9);

$$E(\text{rainfall}) = \beta_1 \text{humidity} + \beta_2 \text{ windspeed} + f(\text{ temperature }) \qquad \text{(Eq.9)}$$

where $f$ is cubic spline approximation. The proposed model for the Lognormal distribution was given as below

$$E(\log[\text{rainfall}]) = f_1(\text{humidity}) + \beta_1 \text{windspeed} + f_2(\text{temperature}) \qquad \text{(Eq.10)}$$


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1404_635644

Hence, the model structure in (9) becomes a combination of linear component of wind speed and humidity and smooth functions of temperature. However model (Eq.10) consists of linear component of wind speed and smooth functions of humidity and temperature.

In *Table 5*, AIC value for the lognormal model was calculated as 1882 whereas the same value for the Gumbel model was 1889. Since Lognormal model provides lower AIC we would use Lognormal model (Eq.10) for the future predictions.

***Table 5***. *Model checking*

| | **Lognormal Model** | **Gumbel Model** |
|---|---|---|
| GD | 1860 | 1873 |
| AIC | 1882 | 1889 |

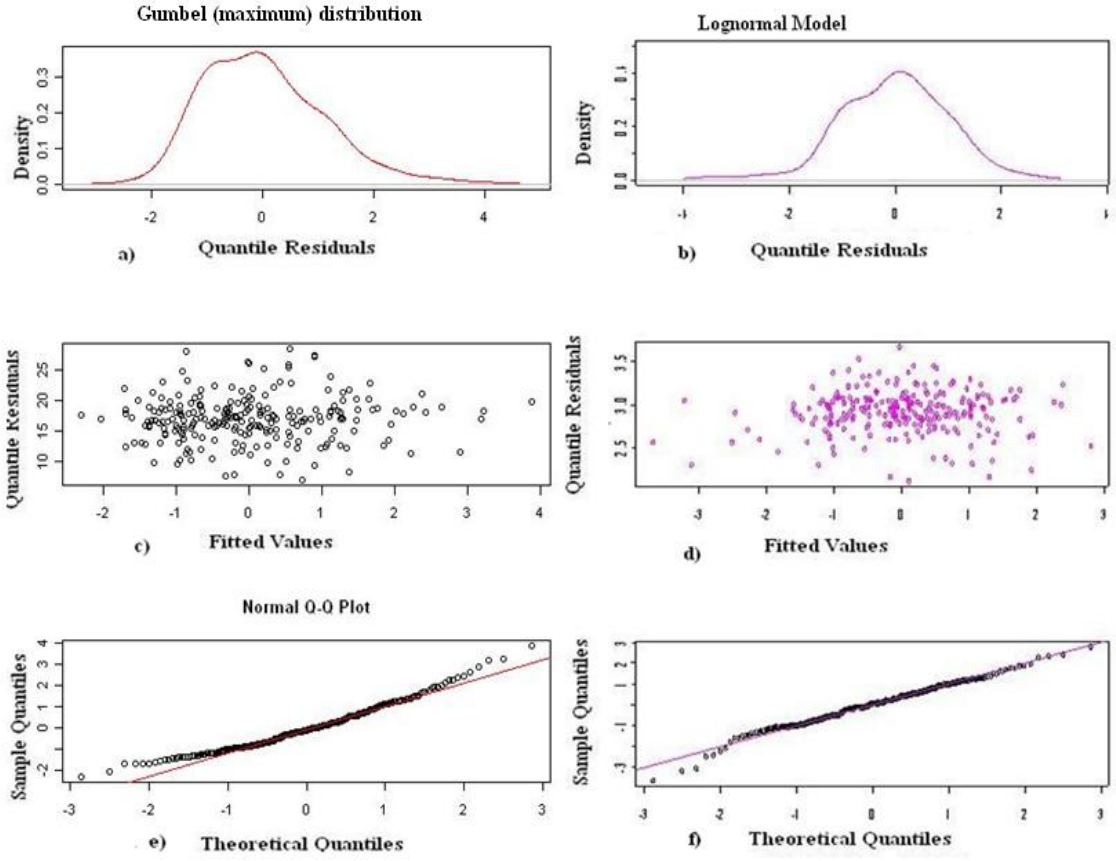


***Figure 3.*** *Summary plots of both model*

For the appropriate fit, the standardized quantile residuals should be close to the standard normal distribution. The quantile residuals of the lognormal model provide a closer distribution to the standard normal distribution than the Gumbel model in *Figure 3.a*. Quantile residuals versus fitted values of both model evenly scattered around zero given in *Figure 3.c-d.* However the Gumbel model shows some departures from the


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1404_635644

theoretical distribution as in *Figure 3.e*. Accordingly, Q-Q plot of lognormal model in *Figure 3.f* provides a better fit than the Gumbel model in *Figure 3.e*. AIC scores and the analysis of quantile residuals together suggest better fit for the lognormal model(10) than the Gumbel model.

## Discussion

In this study, Lognormal model(10) and Gumbel model(9) are fitted to the extreme rainfall data. Although extreme value theory states that there are only three types of distributions are needed to model extreme events (Gumbel, Frechet and Weibull), we identify the model that best characterizes the behavior of the rainfall data is the lognormal model.

Simulation results indicate that even for small sample approximated mean function by the GAMLSS converges to the true function. Moreover, the bias is decreasing rapidly for the true fixed parameter. Although our statistical modeling procedure utilizes rainfall data, the same methodology can be applied to the other environmental variables.

## REFERENCES

[1] Chavez-Demoulin, V., Davison, A. C. (2005): Generalized additive modelling of sample extremes. - Journal of the Royal Statistical Society Series C-Applied Statistics 54:207-222.

[2] Coles, S. (2001): An introduction to statistical modeling of extreme values. - London, New York: Springer.

[3] Coles, S., Pericchi, L. R., Sisson, S. (2003): A fully probabilistic approach to extreme rainfall modeling. Journal of Hydrology 273(1-4), 35-50.

[4] Davison, A. C., Ramesh, N. I. (2000): Local likelihood smoothing of sample extremes. - Journal of the Royal Statistical Society Series B-Statistical Methodology 62: 191-208.

[5] Fisher, R. A., Tippett, L. C. H. (1928): Limiting forms of the frequency distribution of the largest or smallest member of a sample. - Mathematical Proceedings of the Cambridge Philosophical Society 24: 180-190. doi: 10.1017/S0305004100015681

[6] Friedman, J. H., Stuetzle, W. (1981): Projection Pursuit Regression. - Journal of the American Statistical Association 76(376): 817-823.

[7] Gnedenko, B. V. (1943): Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire. - Annals of Mathematics 44(3): 423-453.

[8] Hastie, T., Tibshirani, R. (1987): Generalized Additive Models: Some Applications. [journal]. - Journal of the American Statistical Association 87(398): 371-386.

[9] Hosking, J. R. M., Wallis, J. R. (1997): Regional frequency analysis : an approach based on L-moments. - Cambridge ; New York: Cambridge University Press.

[10] Katz, R., Parlange, M., Naveau, P. (2002) Statistics of extremes in hydrology. - Adv Water Resour 25:1287–1304.

[11] Kharin, V. V., Zwiers, F. W. (2000): Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere-ocean GCM. - Journal of Climate 13(21): 3760-3788.

[12] Lehmann, E. A., Phatak, A., Stephenson, A., Lau, R. (2016): Spatial modelling framework for the characterisation of rainfall extremes at different durations and under climate change. - Environmetrics 27(4): 239-251.

[13] Markose, S., Alentorn, A. (2005): The Generalized Extreme Value (GEV) Distribution, Implied Tail Index and Option Pricing, Discussion Papers. - University of Essex, Department of Economics.


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1404_635644

[14] Rigby, R. A., Stasinopoulos, D. M. (2001): The GAMLSS project: a flexible approach to statistical modelling. - New Trends in Statistical Modelling: Proceedings of the 16th International Workshop on Statistical Modelling, pp. 337-345.
[15] Rigby, R. A., Stasinopoulos, D. M. (2005): Generalized additive models for location, scale and shape. - Journal of the Royal Statistical Society Series C-Applied Statistics 54:507-544.
[16] Ruppert, D., Wand, M. P., Carroll, R. J. (2003): Semiparametric regression. - Cambridge ; New York: Cambridge University Press.
[17] Sang, H. Y., Gelfand, A. E. (2009): Hierarchical modeling for extreme values observed over space and time. Environmental and Ecological Statistics 16(3):407-426. doi: DOI 10.1007/s10651-007-0078-0
[18] Yatchew, A. (2003): Semiparametric regression for the applied econometrician. - New York: Cambridge University Press.
[19] Yee, T. W., Stephenson, A. G. (2007): Vector generalized linear and additive extreme value models. Extremes 10(1-2): 1-19. doi: 10.1007/s10687-007-0032-4

## APPENDIX

**A.1.** Bump function is defined as

$$f(x) = \frac{1}{0.1 + x} + 8\exp(-400(x - 0.5)^2) \qquad \text{for } x \in [0,1].$$

**A.2.** The cyclic function is defined as

$$f(x) = \cos(4\pi x) + 2x \qquad \text{for } x \in [0,1].$$

**A.3.** The probability density function of the Gumbel (maximum) distribution is defined as

$$f_Y(y|\mu,\sigma) = \frac{1}{\sigma}\exp\left[(-\frac{y-\mu}{\sigma}) - \exp(-\frac{y-\mu}{\sigma})\right] \qquad \text{for } -\infty < y < \infty,\ -\infty < \mu < \infty,\ \sigma > 0.$$

The probability density function of the Lognormal distribution is defined as

$$f_Y(y|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\frac{1}{y}\exp\left[-\frac{(\log y - \mu)^2}{2\sigma^2}\right] \qquad \text{for } y > 0,\ \mu > 0,\ \sigma > 0.$$


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1404_635644