DOI: 10.24996/ijs.2023.64.11.27

ISSN: 0067-2904

Modeling Extreme COVID-19 Data in Iraq

Tasnim Hasan Kadhim Albaldawi[1]*, Saleemah H. Jasaim[2]
[1]*Department of Math., College of Science, University of Baghdad, Baghdad, Iraq*
[2]*Department of Astronomy and Space, College of Science, University of Baghdad, Baghdad, Iraq*

Received: 22/11/2022 Accepted: 19/1/2023 Published: 30/11/2023

Abstract

This paper considers the maximum number of weekly cases and deaths caused by the COVID-19 pandemic in Iraq from its outbreak in February 2020 until the first of July 2022. Some probability distributions were fitted to the data. Maximum likelihood estimates were obtained and the goodness of fit tests were performed. Results revealed that the maximum weekly cases were best fitted by the Dagum distribution, which was accepted by three goodness of fit tests. The generalized Pareto distribution best fitted the maximum weekly deaths, which was also accepted by the goodness of fit tests. The statistical analysis was carried out using the Easy-Fit software and Microsoft Excel 2019.

Keywords: Dagum, Generalized Pareto, Generalized Extreme Value, Burr XII, Pearson Type 6

نمذجة البيانات المتطرفة لكوفيد-19 في العراق

تسنيم حسن كاظم البلداوي[1]*، سليمة حمادي جاسم[2]
[1]قسم الرياضيات، كلية العلوم، جامعة بغداد، بغداد، العراق
[2]قسم الفلك والفضاء، كلية العلوم، جامعة بغداد، بغداد، العراق

الخلاصة

هذا البحث يتعامل مع بيانات القيمة المتطرفة عن طريق الاخذ بالاعتبارالقيم العظمى الاسبوعية لحالات الاصابة والوفاة في العراق الناتجة عن جائحة كوفيد-19 منذ تاريخ تفشيها في شباط 2020 ولغاية الاول من تموز 2022. تمت موائمة بعض التوزيعات الاحتمالية.وتم استخراج تقديرات الارجحية العظمى واداء اختبارات جودة الموائمة. اظهرت النتائج ان بيانات القيم العظمى الاسبوعية للاصابات قد توائمت مع توزيع Dagum كأفضل نموذج وكانت اختبارات جودة الموائمة مقبولة النتائج في ثلاثة اختبارات. أما بيانات القيم العظمى الاسبوعية للوفيات فقد توائمت مع توزيع باريتو المعمم كأفضل نموذج وكانت بعض نتائج اختبارات الموائمة مقبولة.

1. Introduction

Since the outbreak of COVID-19 at the end of 2019, it rapidly spread worldwide and causes the death of people of all ages, mostly the elderly and those with certain chronic illnesses [1]. It is of interest to provide the best description of the COVID-19 data by fitting a probability model. Datta, R. *et al* [2] studied the growth of COVID-19 cases using the

*Email: tasnim.h@sc.uobaghdad.edu.iq

structural break methodology. They concluded that the lognormal distribution provides a better fitting to the number of days and the number of cases in each break interval of COVID-19 data in the whole world. Aadhityaa M. *et al* [3] used a statistical modeling approach to study the applicability of extreme value distributions in predicting COVID-19 confirmed cases. Balli S. [4] examined the increase of Covid-19 weekly cases in the USA, Germany, and the globe. He found that the largest extreme value distribution was due for the global and Germany and that the smallest extreme value distribution was due for the USA. Daouia A. *et al* [5] studied the transmission of SARS-CoV-2 using extreme value methods. They suggested fitting the discrete generalized Pareto distribution to the exceedances over a high threshold. In an unpublished paper [6], a study on modeling COVID-19 daily cases and deaths data in Iraq concluded that the extreme value distribution and the generalized Pareto distribution were the best for fitting these data. It was suggested to carry out the analysis with extreme-value data rather than the original data.

This paper considers the maximum number of weekly cases and deaths caused by the COVID-19 pandemic in Iraq from its outbreak in February 2020 until the first of July 2022. Dagum, Burr Type XII, the generalized extreme value, the generalized Pareto, and Pearson Type 6 distributions are proposed for modeling the maximum weekly cases and deaths data. The performance of these distributions is ranked by three goodness of fit tests: the Kolmogorov-Smirnov, Anderson-Darling, and chi-square tests.

2. Statistical Distributions

We will review the statistical distributions used in the study. These distributions are the best fitted to our data.

2.1 Dagum Distribution

The Dagum distribution is proposed by Dagum (1977) [7]. It has many important applications in economics, such as modeling income data. It is further used in models of wealth distribution [8]. Among other applications, it is used in modeling lifetime data. In the actuarial literature, the Dagum distribution is known as the inverse Burr distribution [9]. The three-parameter Dagum distribution pdf is given as

$$f(x) = \frac{\alpha k \left(\frac{x}{\beta}\right)^{\alpha k-1}}{\beta \left(1+\left(\frac{x}{\beta}\right)^{\alpha}\right)^{k+1}}, x > 0;\ k, \alpha, \beta > 0$$

where β is a scale parameter, and α and k are shape parameters. The Dagum distribution is a Burr type III distribution after involving the additional scale parameter β. The Dagum distribution with the shape parameter $k = 1$ is known as the log-logistic distribution. The cumulative distribution function is

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

2.2 Burr Type XII Distribution

The Burr type XII distribution was first introduced by Burr (1942) [10]. This distribution can fit a wide range of data in various fields and has a lot of application areas, especially in quality control, reliability, and environmental data [11]. The Burr type XII distribution is flexible enough to include various shapes and it is among the distributions that are sensitive to extreme values [12]. The probability density function is

$$f(x) = \frac{\alpha k \left(\frac{x}{\beta}\right)^{\alpha-1}}{\beta\left(1+\left(\frac{x}{\beta}\right)^{\alpha}\right)^{k+1}}, x > 0;\ k, \alpha, \beta > 0$$

where β is the scale parameter, α and k are shape parameters. The corresponding distribution function is

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

2.3 The Generalized Extreme Value Distribution

The generalized extreme value distribution is a generalized form of a family of three types of extreme value distributions known as type I (Gumbel), type II (Frechet), and type III (Weibull) extreme value distributions [13]. In other words, the generalized extreme value distribution includes these three types of distributions as special cases [14]. It is used to simulate extreme events and natural disasters [15]. The probability density function is

$$f(x) = \begin{cases} \frac{1}{\sigma}\ exp\left(-(1+kz)^{-1/k}\right)(1+kz)^{-1-1/k} & , k \neq 0 \\ \frac{1}{\sigma}\exp(-z-\exp(-z)) & , k = 0 \end{cases}$$

where $z = \frac{x-\mu}{\sigma}$, and k, $\sigma > 0$, and μ are the shape, scale and location parameters, respectively.

The corresponding cumulative distribution function is

$$F(x) = \begin{cases} exp\left(-(1+kz)^{-1/k}\right), & k \neq 0 \\ \exp(-\exp(-z)), & k = 0 \end{cases}$$

2.4 The Generalized Pareto Distribution

The generalized Pareto distribution was suggested by Pikands (1975) [16] and has been studied by Davison and Smith (1990) [17], Castillo and Hadi (1997) [18] and others. It is widely used as a probability distribution to analyze extreme events that exceed a threshold. The probability density function is

$$f(x) = \begin{cases} \frac{1}{\sigma}\left(1+k\frac{(x-\mu)}{\sigma}\right)^{-1-1/k}, & k \neq 0 \\ \frac{1}{\sigma} exp\left(-\frac{(x-\mu)}{\sigma}\right), & k = 0 \end{cases}$$

The cumulative distribution function is

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

2.5 Pearson Type 6 Distribution

The Pearson family of distributions was proposed by Karl Pearson in 1894 [19]. It is a possible candidate for fitting very large data sets in comparison with other distributions. The beta- and F-distributions are particular cases of this distribution [20]. The probability density function is

$$f(x) = \frac{(x/\beta)^{\alpha_1-1}}{\beta B(\alpha_1,\alpha_2)(1+x/\beta)^{\alpha_1+\alpha_2}}, x > 0;\ \alpha_1, \alpha_2, \beta > 0$$

The cumulative distribution function is
$F(x) = I_{x/(x+\beta)}(\alpha_1, \alpha_2)$
where B is the beta function, and $I_{x/(x+\beta)}(\alpha_1, \alpha_2)$ is the incomplete beta function.

3. Analysis and Results

Maximum weekly confirmed cases and deaths were extracted from daily COVID-19 cases and deaths in Iraq from the 24th of February 2020 to the 1st of July 2022. Firstly, descriptive statistics were presented for the resulting extreme value data. Thereafter, five probability distributions were selected to model the data. These distributions are known to fit such extreme-value data well. The selected distributions are the Dagum, Burr Type XII, generalized extreme value, generalized Pareto, and Pearson Type 6 distributions. Maximum likelihood estimates for the parameters are obtained. The performance of these distributions is ranked by three goodness-of-fit tests; the Kolmogorov-Smirnov, Anderson-Darling, and chi-square tests. This is done using Easy-Fit software.

3.1 Analysis of Maximum Weekly COVID-19 Cases

Table 1 shows the summary statistics and percentile values for the resulting 123 maximum weekly recorded cases. The variance or the standard deviation reveals high variability in the data. The coefficient of variation is close to one since the value of the mean is nearly equal to the value of the standard deviation. Furthermore, the data are positively skewed. A plot of the COVID-19 maximum weekly cases is depicted in Figure 1. We notice that COVID-19 pandemic cases in Iraq have passed through four peaks during the period of the study, and a fifth is proceeding.

Table 1: Descriptive statistics and percentile values of maximum weekly cases

Statistic	No. of Weeks	Range	Mean	Variance	Standard Deviation	Coef. of Variation	Stand. Error	Coef. of Skewness	Coef. of Kurtosis
Value	123	13510	3203.7	9.2348E+6	3038.9	0.94856	274.01	1.0548	0.67545

Percentile	Min.	5%	10%	25% (Q1)	50% Median	75% (Q3)	90%	95%	Max.
Value	5	53.6	118.4	625	2519	4938	8141.0	9117.2	13515

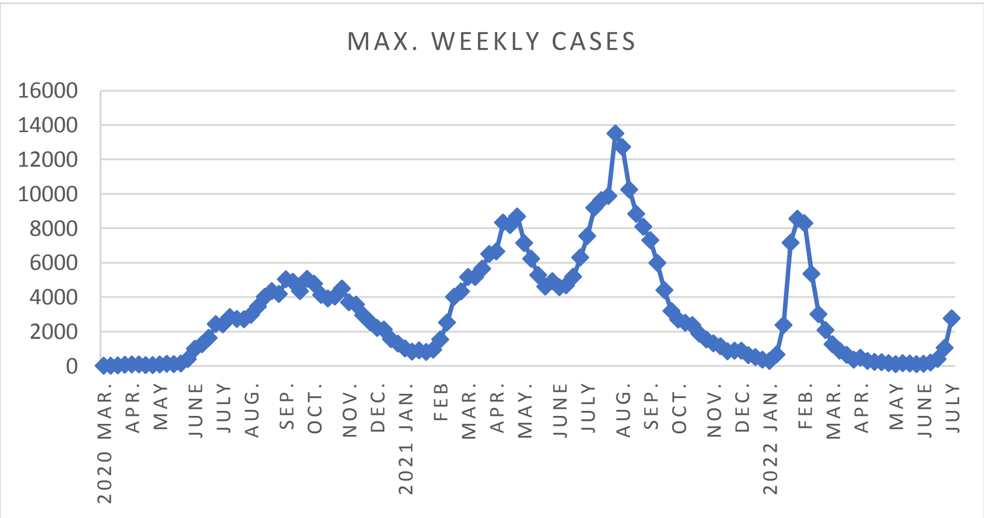

Figure 1: COVID-19 number of the maximum weekly cases

Fitting results are presented in Tables 2 and 3. Table 2 presents the maximum likelihood estimates of the parameters of the five distributions. Table 3 presents the goodness of fit results and the ranking of each model according to each test. It is obvious that the Dagum distribution is the best-fitting distribution. The second best-fit model is the generalized Pareto distribution. Both of these distributions are not rejected by the null hypothesis that the data follow the assigned distribution according to the three goodness of fit tests. A PP plot for the fitted Dagum model of maximum weekly COVID-19 cases is depicted in Figure 2. It sounds like the Dagum model was adequate for fitting our extreme value data, especially for the smallest and largest extremes.

Table 2: Maximum likelihood estimates

Distribution	Parameters
Burr XII	$k = 647.87$ $\alpha = 0.86332$ $\beta = 5.3975E + 6$
Dagum	$k = 0.09253$ $\alpha = 5.9727$ $\beta = 8589.6$
Gen. Extreme Value	$k = 0.12272$ $\sigma = 2101.7$ $\mu = 1702.4$
Gen. Pareto	$k = -0.19677$ $\sigma = 4384.0$ $\mu = -429.39$
Pearson 6	$\alpha_1 = 0.76637$ $\alpha_2 = 1.1430E + 7$ $\beta = 4.6785E + 10$

Table 3: Distribution ranking and goodness of fit test statistics

Distribution	Kolmogorov Smirnov Critical Value=0.12245 (α=0.05)			Anderson-Darling Critical Value=2.5018 (α=0.05)			Chi-Squared Critical Value=12.592 (α=0.05)		
	Statistic	Reject?	Rank	Statistic	Reject?	Rank	Statistic	Reject?	Rank
Burr XII	0.0992	No	3	1.7068	No	4	15.254	Yes	4
Dagum	0.04658	No	1	0.35837	No	1	2.0209	No	1
Gen. Extreme Value	0.10242	No	5	2.0958	No	5	14.706	Yes	3
Gen. Pareto	0.09598	No	2	1.1292	No	2	4.6838	No	2
Pearson 6	0.10191	No	4	1.4654	No	3	15.792	Yes	5

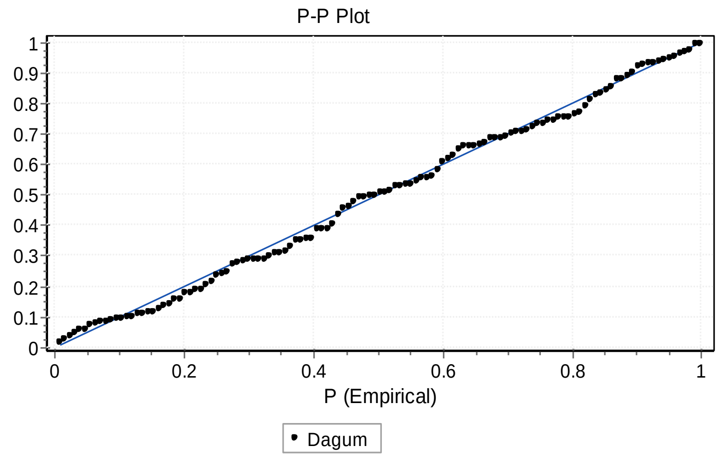

Figure 2: A PP Plot for the best-fitted model of maximum weekly cases

3.2 Analysis of Maximum Weekly COVID-19 Deaths

Table 4 shows the summary statistics and percentile values for the maximum weekly confirmed deaths. The data are somehow less dispersed, with a lower coefficient of variation and less positive skewness than that assigned in Table 1. Figure 3 depicts the maximum number of weekly deaths during the study period. We can notice the high peak of deaths at the beginning of the pandemic, followed by two lower peaks. The substantial reduction in the number of deaths is due to the enforcement of taking the vaccine at the end of the study period.

Table 4: Descriptive statistics and percentile values of maximum weekly deaths

Statistic	No. of Weeks	Range	Mean	Variance	Standard Deviation	Coef. of Variation	Stand. Error	Coef. of Skewness	Coef. of Kurtosis
Value	123	122	36.024	971.38	31.167	0.86517	2.8102	0.79097	-0.31827

Percentile	Min.	5%	10%	25% (Q1)	50% Median	75% (Q3)	90%	95%	Max
Value	0	1	2	8	31	55	84	93.6	122

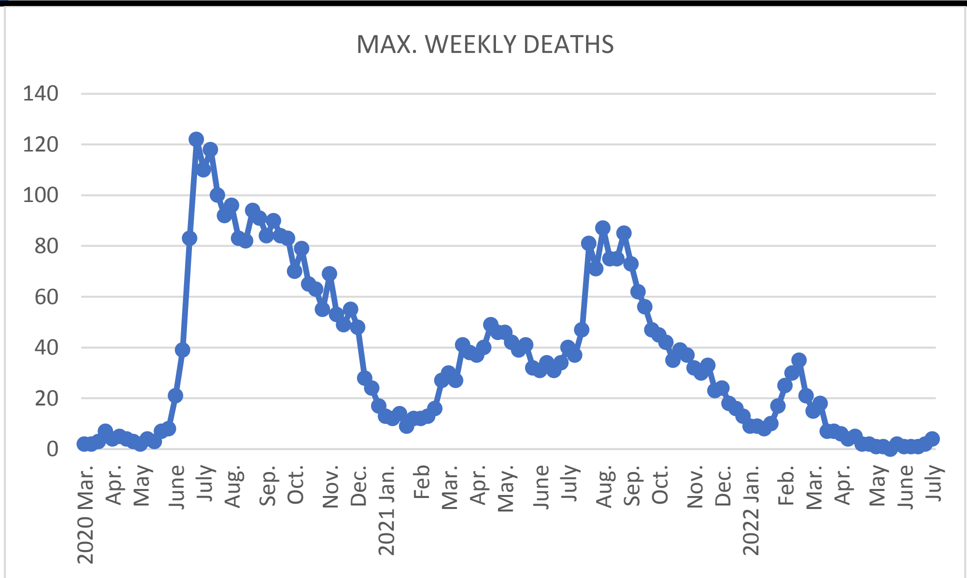

Figure 3: COVID-19 number of the maximum weekly deaths

The results of the maximum weekly number of deaths are shown in Tables 5 and 6. Table 5 shows the maximum likelihood estimates of the parameters. Table 6 shows the goodness of fit results and the ranking of each distribution according to each test. It can be seen that the generalized Pareto distribution has the highest rank according to the three goodness of fit tests. However, this distribution is not rejected by the null hypothesis that the data fit the assigned distribution.

Taking into account the peak value of 122 deaths, the peak over threshold approach seems to work well for our extreme value data. A large number of studies indicate that it is the most effective approach for fitting the extreme value data [21]. Figure 4 depicts a PP plot of the fitted generalized Pareto model of maximum weekly COVID-19 deaths.

Table 5: Maximum likelihood estimates

Distribution	Parameters
Burr XII	$k = 1.5281E + 5$ $\alpha = 1.0412$ $\beta = 3.5411E + 6$
Dagum	$k = 0.06037$ $\alpha = 10.628$ $\beta = 97.396$
Gen. Extreme Value	$k = 0.06372$ $\sigma = 23.495$ $\mu = 20.889$
Gen. Pareto	$k = -0.3016$ $\sigma = 51.928$ $\mu = -3.8707$
Pearson 6	$\alpha_1 = 1.0083$ $\alpha_2 = 5.5607E + 6$ $\beta = 2.0202E + 8$

Table 6: Distribution ranking and goodness of fit test statistics

Distribution	Kolmogorov Smirnov Critical Value=0.12263 (α=0.05)			Anderson-Darling Critical Value=2.5018 (α=0.05)			Chi-Squared Critical Value=12.592 (α=0.05)		
	Statistic	Reject?	Rank	Statistic	Reject?	Rank	Statistic	Reject?	Rank
Burr XII	0.08535	No	2	5.4849	Yes	5	10.23	No	3
Dagum	0.09631	No	5	4.7737	Yes	3	8.7074	No	2
Gen. Extreme Value	0.08867	No	4	1.7638	No	2	13.589	Yes	4
Gen. Pareto	0.0745	No	1	0.87997	No	1	5.765	No	1
Pearson 6	0.08671	No	3	5.2586	Yes	4	14.636	Yes	5

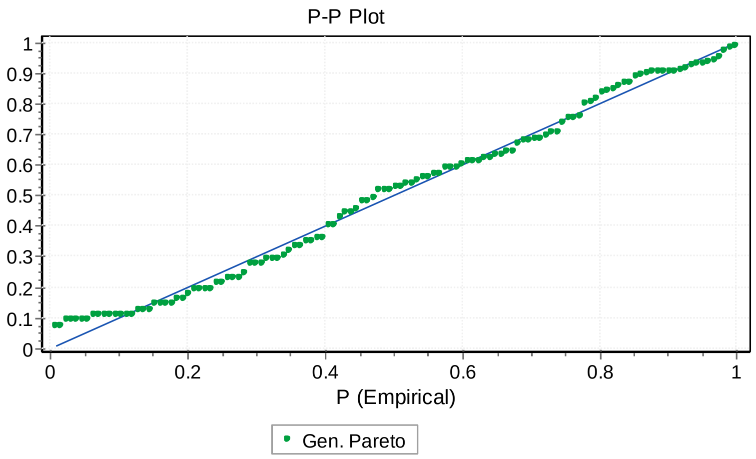

Figure 4: A PP Plot for the best-fitted model of maximum weekly deaths

4. Conclusions

It was found that the Dagum model fits the maximum weekly COVID-19 cases quite well. Further, it was also found that the generalized Pareto distribution is the best model for fitting the maximum weekly COVID-19 deaths. Finally, we conclude that in dealing with extreme events like the COVID-19 pandemic, working with extreme-value data was the right choice.

References

[1] WHO, 2020. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen.

[2] R. Datta, P. Trivedi, A. Kumawat, R. Kumar, I. Bhardwaj, N. Kumari, V. Agiwal, S. Kumar, A. Kumar, A. Shukla and J. Kumar, "Statistical Modeling of COVID-19 Pandemic Stages Worldwide," Preprints 2020, doi: 10.20944/preprints 202005.0319.v1.

[3] K. S. Aadhityaa, I. Kasiviswanathan, B. Ilampooranan and J. H. Soundharajan, "A Global Scale Estimate of Novel Coronavirus (COVID-19) Cases Using Extreme Value Distributions," medRxiv doi: https://doi.org/10.1101/2020.04.17.20069500, 2020.

[4] S. Balli, "Data analysis of Covid-19 pandemic and short-term cumulative case forecasting using machine learning time series methods," *Chaos, Solitons and Fractals*, vol. 142, no. 110512, 2021.

[5] A. Daouia, S. Gilles and A. Usseglio-Carleve, "Extreme Value Modelling of SARS-CoV-2 Community Transmission Using Discrete Generalised Pareto Distributions," ffhal-03392044v2., 2022.

[6] T. H. K. AlBaldawi, "Statistical Analysis of COVID-19 Data in Iraq," Accepted paper to be published at the *Iraqi Journal of Science*, vol. 64, no. 9, 2023.

[7] C. A. Dagum, "A New Model of Personal Income Distribution: Specification and Estimation," *Economie Appliqu'ee*, vol. 30, pp. 413-437, 1977.

[8] J. G. Vargas and D. V. Voiculescu, "Boolean Extremes and Dagum Distributions," arXiv:1711.06227v2 [math. FA], 2018.

[9] C. Kleiber, "A guide to the Dagum distributions," in Duangkamon Chotikapanich (ed.) Modelling income distributions and Lorenz curves. (Economic Studies in Inequality, Social Exclusion and Well-Being), chapter 6, Springer, 2008.

[10] I. W. Burr, "Cumulative frequency functions," Annals of Mathematical Statistics, vol. 13, no. 2, pp. 215-232 doi: 10.1214/aoms/1177731607, JSTOR 2235756, 1942.

[11] W. M. Thupeng, "Use of the Three-parameter Burr XII Distribution for Modelling Ambient Daily Maximum Nitrogen Dioxide Concentrations in the Gaborone Fire Brigade," *American Academic Scientific Research Journal for Engineering, Technology, and Sciences*, vol. 26, no. 2, 2016.

[12] A. Y. AL-Saiari, L. A. Baharith and S. A. Mousa, "Marshal Olkin extended Burr type XII distribution," *International Journal of Statistics and probability*, vol. 3, no. 1, pp. 78-84, 2014.

[13] T. H. K. Albaldawi and Z. Z. A. Alzuabidi, "Extreme Value Analysis of Maximum Rainfull Data in Baghdad City," *Mathematics and Statistics Journal*, vol. 2, no. 3, pp. 1-8, 2016.

[14] T. H. K. Albaldawi and Z. Z. A. Alzuabidi, "Statistical Analysis of Extreme Rainfall Data in Baghdad City," *Iraqi Journal of Science*, vol. 57, no. 1C, pp. 713-718, 2016.

[15] F. N. Ogana, J. S. A. Osho and J. J. Gorgoso-Varela, "Application of extreme value distribution for assigning optimum fractions to distributions with boundary parameters: an eucalyptus plantations case study," *Sib. J. for. Sci*., no. 4, pp. 39-48, 2018.

[16] J. Pickands, "Statistical inference using extreme order statistics," *Ann. Stat*., vol. 3, pp. 119-130, 1975.

[17] A. Davidson and R. Smith, "Models for exceedances over high thresholds (with discussion)," *JR Stat Soc*., vol. 52, pp. 393-442, 1990.

[18] E. Castillo and A. Hadi, "Fitting the Generalized Pareto Distribution to Data," *Journal of the American Statistical Association*, vol. 92, pp. 1609-1620, 1997.

[19] K. Pearson, "Contributions to the mathematical theory of evolution, II: Skew variation in homogeneous material," *Philos. Trans. Royal Soc. London*, ARRAY, pp. 343-414, 1895.

[20] B. Lahcene, "On Pearson families of distributions and its applications," *African Journal of Mathematics and Computer Science Research*, vol. 6, no. 5, pp. 108-117, 2013 DOI: 10.5897/AJMCSR2013.0465.

[21] M. Khan, F. Aslam and P. Ferreira, "Extreme Value Theory and COVID-19 Pandemic: Evidence from India," *The Economic Research Guardian*, vol. 11, no. 1, pp. 2-10, 2021.