

Original Article

# Modelling Daily Rainfall Amount in Pekanbaru City using Gamma and Some Extended Gamma Distribution

Muhammad Rajab<sup>1</sup>, Rado Yendra<sup>2</sup>, Muhammad Marizal<sup>3</sup>, Ari Pani Desvina<sup>4</sup>, Rahmadeni<sup>5</sup>

<sup>1,2,3,4,5</sup> Department of Mathematics, Faculty of Science and Technology, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, 28293, Indonesia.

Received: 06 April 2022

Revised: 23 May 2022

Accepted: 07 June 2022

Published: 24 June 2022

**Abstract** - Modeling rainfall is very important to be developed in managing natural resources to deal with the impacts of climate change. We modelled the daily rainfall for data recorded in Pekanbaru City from 1999 to 2008. the main goal of this study is to find the best fitting distribution to the daily rainfalls by using the maximum likelihood approach. for this purpose, Gamma distribution and some Extended Gamma Distribution will be used and tested to determine the best model to describe daily rainfall in Pekanbaru City. the extended gamma distribution meaning some mixture two and three gamma distribution, namely rani, shanker and sujatha distribution. the maximum likelihood method will be used to get the estimated parameter value from the distribution used in this study. the distributions will be selected based on graphical inspection probability density function (pdf), numerical criteria Akaike's information criterion (AIC) and Bayesian Information Criterion (BIC). in most the cases, graphical inspection gave the same result but their AIC and BIC result differed. the best fit result was chosen as the distribution with the lowest values of AIC and BIC. in general, the Gamma distribution has been selected as the best model

**Keywords** - Ektented Gamma Distribution, Rainfall Modelling, Rani Distribution, Shanker Distribution, Sujatha Distribution.

## 1. Introduction

Clusterwise regression is a method that finds simultaneously an optimal member of data in k cluster and each cluster have the best regression model. Analysis of a simulated data set has also been presented for illustrative purposes. Gamma and normal distributions were used for distribution of responses scenario with different parameters. This simulation study is carried out by initializing the number of clusters, classify observations randomly as an initial partition, move observation to the cluster giving the smallest residual and re-estimate the regression model from final partition. This simulation showed that clusterwise regression is able to form partition according to the distribution of data, also to form the best generalized linear model with Gamma distribution and linear regression model [1].

There are many observations in the economics, social and science that arise with a high variance. When the data was estimated a single regression model there was a mistake in presenting the data structure. the data can be classified to reduce variance. One of the ways to classify the data is clusterwise regression techniques. the clusterwise regression is based on the combination of these two techniques that find simultaneously an optimal partition of data in k cluster and regression function within cluster [2]. It is assumed that samples come from a certain number of populations and consider the existence of subpopulations of heterogeneous populations. the proportion of subpopulations is unknown. A specific form for each subpopulation can be determined and the purpose of clusterwise regression is to describe the sample into mixture components based on the subpopulation. Estimating parameters in the cluster wise regression method is needed to estimate the regression coefficients for each cluster. Various kinds of algorithms are formed to overcome this problem, among others, based on exchange algorithm [3], statistical techniques [4], and optimization techniques [5] [6]. the application of cluster regression to various fields of research has been carried out, such as business research, physics, and social studies. Data not only arise from normal distribution but also arise from exponential families. Statistical techniques from [4] developed the cluster wise method for the approach for generalized linear models.

Water resource assessment involved various variables that can be simplified and tackled by developing a suitable mathematical model. Rainfall-Runoff (RR) modeling considered as a major hydrologic process and is essential for water resources management. This study presents the development of rainfall-runoff model based on artificial neural networks



(ANNs) models in Shipra river basin of Madhya Pradesh. the ability of model was evaluated based on sum of squares error (SSE) and relative error. the Sum of squares error obtained during this study was 30.525 in training and 53.076 in testing and the Relative error value obtained was 0.939 in training and 0.874 in testing at Mahidpur station but at Ujjain station, the SSE obtained during this study was found to be 30.488 during training and 10.703 during testing while the relative error value obtained was 0.938 in training and 0.915 in testing. the model was found suitable for simulating hydrological response of the basin to the rainfall and predicting daily runoff with high degree of accuracy. the study demonstrates the applicability of ANN approach using the statistical tool SPSS 16.0 in developing effective non-linear models of rainfall-runoff process in order to represent the internal hydrologic structure of the watershed [7]. We need to study the basin response to the catchment rainfall for water resource planning of a basin. This requires development of a relationship between basin rainfall and runoff. Most of river catchments in India are ungauged and generally the limited discharge data are available with the concern state and central agencies. Under such circumstance's rainfall-runoff model can be developed to simulate the natural hydrological processes to estimate the runoff from the catchment. A rainfall-runoff model is a mathematical. representation describing the rainfall-runoff relations of a catchment area or a watershed. More precisely, it produces the surface runoff hydrograph as a response to a rainfall as an input. Rainfall-runoff models are classified as deterministic, stochastic, conceptual, theoretical, black box, continuous event, complete, routing or simplified [8]. the widely known rainfall-runoff models identified are the rational method [9] Soil Conservation Services (SCS) Curve Number method [10] and Green-Ampt method [11]. Considered watershed as a series of identical reservoirs and prepared a conceptual rainfallrunoff models by routing a unit inflow through the reservoirs [12]. Kumbhare and Rastogi (1984) tested the Nash conceptual model (1958) and found that runoff was generated in good agreement with actual runoff hydrograph [13]. Kumar and Rastogi (1989) developed a mathematical model of the instantaneous unit hydrograph based on time area histogram for a small watershed at Pantnagar [14]. Now-a-days, artificial neural networks (ANN) have found increasing applications in various aspects of hydrology. ANN approach is faster compared with its conventional compatriots, robust in noisy environments, flexible in the range of problems it can solve, and highly adaptive to the newer environments. Data-driven black box models such as ANNs are preferred alternatives for systems in which different mechanisms impact each other and precise identification of the interactions among all these mechanisms is not possible.

The expanded generalization of both function generalization and gamma distribution has been highlighted in recent decades for their widespread use in important application areas in pure and mixture models, and this expansion provides a comprehensive mathematical treatment to trace the behavior of some random phenomena represented by this distribution. the author introduces some key characteristics of extended generalization to gamma, modified and extended gamma, and provides moments, quantitative function, and other important metrics. the findings of this work will be useful and constructive for practitioners in various fields of theoretical and applied sciences [15]. the gamma distribution is the most popular model for analysing skewed data and hydrological processes. Standard lifetime distributions usually present very strong restrictions to produce bathtub curves, and thus appear to be inappropriate for analysing data with this characteristic. the three-parameter generalized gamma distribution [18] includes as special models the exponential, Weibull, gamma and Rayleigh distributions, among others [16, 17, 18]. One of the important families of distributions in lifetime tests is the extended gamma distribution. the extended generalized gamma distribution is a highly known distribution due to its utility in modelling lifetime data where the hazard rate function is monotone in special cases [19, 20]. in 2004, Bachioua proposed an extended gamma distribution with the special feature of being able to adapt failure rates with bathtub and unimodal shape curves [21].

Extended Generalized gamma distribution has proved to be of considerable interest in the field of reliability. It is a reasonable model for life-time distribution of a component (or a system of components) [22, 23]. This distribution does not have the expected impact due to its complicated form. Among these, the extended five gamma distribution was presented by Bachioua et al., among others. A good review of these models is presented in Abdul Moiz and Bachioua [23]. Large-scale rainfall data such as daily is a useful input in statistical models to produce simulations of rainfall data for several years to come as an effort to maintain the availability of water resources. Modelling of daily rainfall data using some statistical modelling have been done to understanding about the rainfall pattern and its characteristics. the statistical modelling of rainfall has become well established over the past 30 years. in particular, the gamma distribution has been used many times to model rainfall totals on wet days [24]. Although many daily rainfall modelling studies use a single distribution, there are several studies that use a mixed distribution. Daily rainfall modeling is very well done using a mixture of exponential distributions [25], [26]. Single log normal distribution with two parameters is not good when compared to a mixture of two log normal distributions in modeling daily rainfall [27]. Shimizu [28] suggests using a mixed log normal distribution as a daily rainfall probability model if there are zero data. Selection of the best distribution in modeling daily rainfall is a major part of researching rainfall. in this study we will determine the best distribution for daily rain based on several goodness-of-fit tests namely on graphical inspection probability density function (PDF) and numerical criteria (AIC and BIC). the objective

of this study is to propose Gamma Distributions and some extended Gamma Distributions namely Rani Distribution or Two mixture of Gamma (1,  $\beta$ ) and Gamma (5,  $\beta$ ) with their weighted  $\frac{\beta^5}{\beta^5+24}$  and  $\frac{24}{\beta^5+24}$  respectively [29], Shanker Distribution or Two mixture of Gamma (1,  $\beta$ ) and Gamma (2,  $\beta$ ) with their weighted  $\frac{\beta^2}{\beta^2+1}$  and  $\frac{1}{\beta^2+1}$  respectively [28] and Sujatha Distribution or Three mixture of Gamma(1,  $\beta$ ), Gamma(2,  $\beta$ ) and Gamma(3,  $\beta$ ) with their mixing weighted  $\frac{\beta^2}{\beta^2+\beta+2}$ ,  $\frac{\beta}{\beta^2+\beta+2}$ ,  $\frac{2}{\beta^2+\beta+2}$  respectively [30]. Comparison of the proposed mixture distributions with existing distribution functions is done to demonstrate their suitability in describing daily rainfall characteristics.

## 2. Data and Study Area

Pekanbaru is a big city which is the capital of Riau Province. Pekanbaru which has a tropical climate, with daily rainfall with varying amount. the initial information about the nature of daily rainfall in Pekanbaru can be seen in table 1. Some basic statistics are displayed based on daily rainfall data in PekanbaruFrom. From Table 1, it can be seen that the variation of rainfall is 304.1024 which means that the city of Pekanbaru has a very varied rainfall from 0.10 m/s to 93.30 m/s. the average daily rainfall of 13.94 m/s indicates that Pekanbaru is still experiencing less heavy rain. .the original data consisted of wind speed records from 1999 to 2008, which were provided by the meteorological, climatological, and geophysical agency of Pekanbaru city, Indonesia. the data and the histogram or characteristic wind speed are presented in Fig. 1. From the histogram data it can also be seen that the rain modeling can use an unbalanced distribution such as the Gamma distribution and which has the same properties as it.

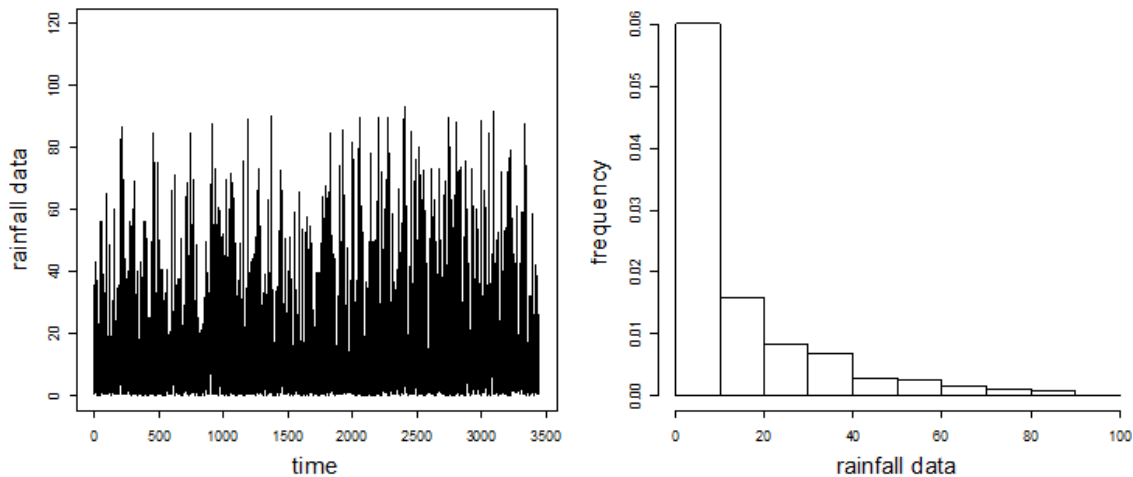


Fig. 1 Plot and histogram daily rainfall data on Pekanbaru respectively

Table 1. The descriptive statistics for daily Rainfall

Statistics	Mean	Variation	Minimum	Maximum
	13.94	304.1024	0.10	93.30

## 3. Methods

### 3.1. Probability Density Function

Rainfall modelling requires analysis of daily rainfall data over a number of years. the primary tools to describe rainfall characteristics are probability density functions. Four probability density function namely the Two parameter gamma distribution with continuous shape parameter ( $\alpha$ ), continuous scale parameter ( $\beta$ ) and some extended Gamma Distributions with continuous scale parameter ( $\beta$ ) (Rani, Shanker and Sujatha) associated with modeling daily rainfall, are considered in this paper. the probability density function for each distribution that we consider are as given in Table 2, where  $y$  denote the observed values of the random variable representing the event of interest.

**Table 2. Probability Density Function (PDF) Gamma and some Extended Gamma distribution**

No	Gamma and some Extended Gamma distribution	Probability density function (PDF) dan distributin function (CDF)
1	Gamma $(\alpha, \beta)^*$	$f(y, \beta) = \frac{\beta}{\Gamma(\alpha)} y^{\alpha-1} \exp(-y\beta), \quad y > 0, \beta > 0$
2	Two mixture of Gamma $(1, \beta)$ and Gamma $(5, \beta)$ with their weighted $\frac{\beta^5}{\beta^5+24}$ and $\frac{24}{\beta^5+24}$ respectively <sup>**)</sup>	$f(y, \beta) = \frac{\beta^5}{\beta^5 + 24} (\beta + y^4) e^{-y\beta}, y > 0, \beta > 0$
3	Two mixture of Gamma $(1, \beta)$ and Gamma $(2, \beta)$ with their weighted $\frac{\beta^2}{\beta^2+1}$ and $\frac{1}{\beta^2+1}$ respectively <sup>***)</sup>	$f(y, \beta) = \frac{\beta^2}{\beta^2 + 1} (\beta + y) e^{-y\beta}, y > 0, \beta > 0$
4	Three mixture of Gamma $(1, \beta)$ , Gamma $(2, \beta)$ and Gamma $(3, \beta)$ with their mixing weighted $\frac{\beta^2}{\beta^2+\beta+2}$ , $\frac{\beta}{\beta^2+\beta+2}$ , $\frac{2}{\beta^2+\beta+2}$ <sup>****)</sup>	$f(y, \beta) = \frac{\beta^3}{\beta^2 + \beta + 2} (1 + y + y^2) e^{-y\beta}, y > 0, \beta > 0$

the Gamma and Some Extended Gamma distribution have known a Gamma distribution<sup>\*</sup>), Rani distribution<sup>\*\*)</sup>, Shanker distribution<sup>\*\*\*)</sup>, and Sujatha Distribution<sup>\*\*\*\*)</sup>

For selecting the best fit model, choice of the model definition, parameter estimation tools are important. the parameter estimation of the distribution function are calculated using maximum likelihood method. the function of maximum likelihood for this model is in implicit and complicated and we will not discuss detail in this paper. the non-linear equation generated by the maximum log likelihood function ( $\ln L$ ) requires a numerical method, namely the newton rhapsion to get the solution of the equation. However this method has been employed in the iteration system to find the solution. Some initial value have been tested for this procedure. If the initial value coverge to the same value, it is considered to be the chosen estimated parameter. the procedure of goodness of fit tests for model selection, both numerically and graphically, is discussed.

**3.2. Mximum Likelihood Estimate (MLE) and Goodness of Fit Tests (GOF)**

Let  $(y_1, y_2, \dots, y_n)$  be random sample from Gamma and Extended Gamma Distributions. the natural log likelihood ( $\ln L$ ) are presented in Table 3. the MLE  $\hat{\theta}$  of  $\theta$  is the solution of the equation  $\frac{d \ln L}{d \theta} = 0$  and thus it is the solution of the following nonlinear equation. the most appropriate distribution is identified using results found based on several goodness-of-fit tests. the GOF tests considered are based on graphical inspection probability density function (PDF) and numerical criteria Akaike's information criterion (AIC) and Bayesian information criterion (BIC) were applied to determine the goodness-of-fit criteria of the distributions. in most the cases, graphical inspection gave the same result but their AIC and BIC result differed. the best fit result was chosen as the distribution with the lowest values of AIC

**Table 3. Log Likelihood function ( $\ln L$ ) for Gamma and Some Extended Gamma Distrbitutons**

Distribution	$\ln L$
Gamma	$n \log(\beta) + (\alpha - 1) \sum \log(x) - \beta \sum x - n \log(\Gamma(\alpha))$
Rani	$n \log\left(\frac{\beta^5}{\beta^5 + 24}\right) + \sum \log(\beta + y^4) - n\beta\bar{y}$
Shanker	$n(\log(\beta^2) - \log(\beta^2 + 1)) + \sum \log(\beta + y) - n\beta\bar{y}$
Sujatha	$n(\log(\beta^3) - \log(\beta^2 + \beta + 2)) + \sum \log(1 + y + y^2) - n\beta\bar{y}$

the formula for computing AIC and BIC are as follows:  $AIC = -2 \ln L + 2k$ ,  $BIC = -2 \ln L + k \ln n$ , where  $k$  = the number of parameters,  $n$  = the sample size

### 4. Result and Discussion

The fitting of daily rainfall collected across Pekanbaru part of Riau Province was considered using data from the period between 1999 and 2008. the data used for are presented and also the daily rainfall data histogram are presented on Figure 1. for the purpose of modelling the rainfall, various distributions have been used, such as Gamma distribution , Rani distribution, Shanker distribution and Sujatha distribution. Behavior of the pdf for varying values of parameters  $\beta$  shown in Fig. 2. It is clearly seen from the graphs of pdf it is obvious that four simple mixture distribution is monotonically decreasing. As the value of parameter  $\beta$  increases.

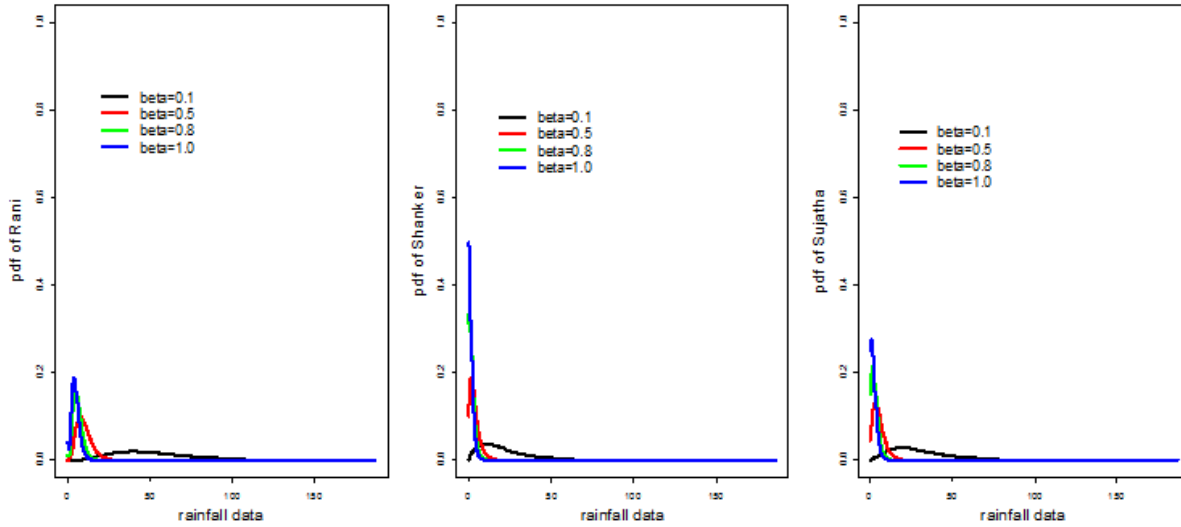


Fig. 2 Graph of pdf some Extended Gamma Distributions (the Rani, Shanker and Sujatha) for different values of the parameter beta ( $\beta$ ) .

Tables 4 shows the estimated parameter value and Table 5 show the statistical parameters for GOF evaluation of pdf namely AIC and BIC currently analyzed for the Gamma distribution and some extended Gamma distributions at the Pekanbaru station. From Table 5, by comparing each model, it is clear that the Rani have the highest AIC values, implying that pdf is not a good model for rainfall in Pekanbaru. However, Gamma distribution has the lowest AIC and BIC values, which implies that Gamma distribution models provides a model that more adequately fits the data.

Table 4. The estimated parameters distributions

	Gamma	Rani	Shanker	Sujatha
$\alpha$ (shape)	0.651	-	-	-
$\beta$ (scale)	0.047	0.369	0.146	0.206

Table 5. The  $\ln L$ , AIC and BIC values for daily rainfall data

	Gamma	Rani	Shanker	Sujatha
$\ln L$	-12259.74	-20651.96	-14191.61	-15300.37
AIC	24523.47	41305.93	28385.21	30602.75
BIC	24535.76	41312.07	28391.35	30608.89

Fig. 3 shows the fitted for Gamma and some Extended Gamma distribution, based on pdf. From this figure Gamma probability density function is very close to the histogram data, this can be interpreted Gamma distribution model is able to provide a good result for daily rainfall data.

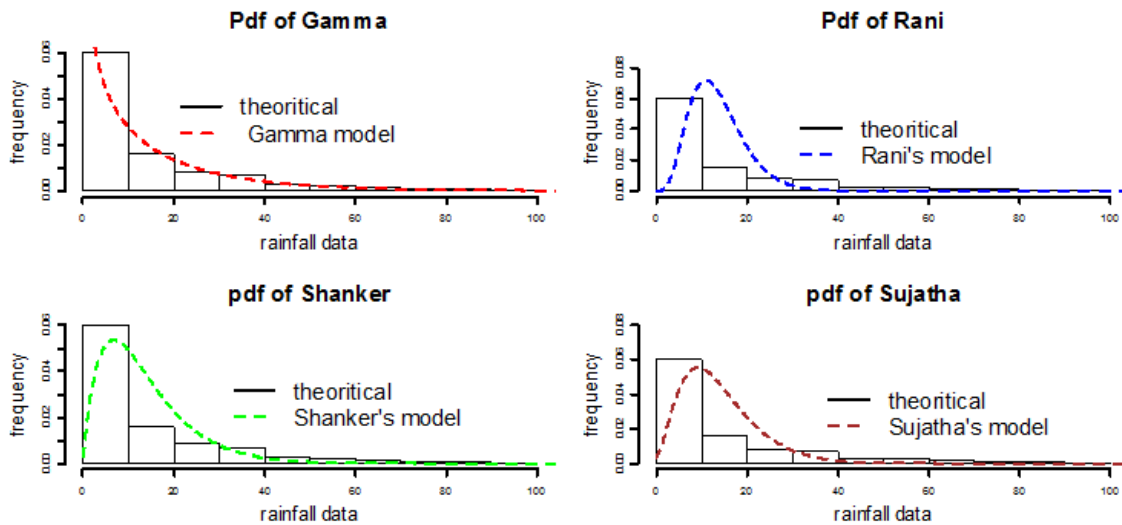


Fig. 3 Fitted pdf plots of Gamma and some Extended Gamma distributions (Rani, Shanker and Sujatha) for the given dataset.

## 5. Conclusion

The search for the best distribution in fitting daily rainfall has become a main interest in several studies. In this research, we focused on determining the best statistical model for daily rainfall in Pekanbaru. The four distributions, namely Gamma and some extended Gamma distributions (Rani, Shanker, and Sujatha), were fitted to the data. The results obtained based on graphical and AIC values indicated that the Gamma distribution adequately modeled the daily rainfall distribution in Pekanbaru. Additionally, from the pdf curve of the Gamma distribution, it can be seen that the most probable daily rainfall occurs at a 0-10 m/s.

## Acknowledgments

The authors are grateful to the Meteorological, Climatological, and Geophysical Agency (BMKG) of Pekanbaru city, Indonesia for supplying data.

## References

- [1] R. Syafruddin, A. M. Soleh, and A. H. Wigena, Clusterwise Regression Model Development with Gamma Distribution. *Icsa*. 1. 2019
- [2] A. M. Bagirov, A. Mahmood, and A. Barton, Prediction of Monthly Rainfall in Victoria, Australia: Clusterwise Linear Regression. *J Atmosres.* (2017) 20-29.
- [3] H. Spath. Algorithm 39: Clusterwise Linear Regression. *Computing*. 22 (1979) 367-373.
- [4] M. Wedel, and W. S. Desarbo. A Mixture Likelihood Approach for Generalized Linear Models. *Journal of Classification*. 12 (1995) 21-55
- [5] A. M. Bagirov, J. Ugon and H. Mirzayeva, An Algorithm for Clusterwise Linear Regression Based on Smoothing Techniques. *Optimization Letters*. 9(1) (2015) 375-390.
- [6] A. M. Bagirov, J. Ugon and H. Mirzayeva, Nonsmooth Nonconvex Optimization Approach to Clusterwise Linear Regression Problems. *Eropan Journal of Operational Research*. 229(1) (2013) 132-142.
- [7] A. K. Yadav, Rainfall-Runoff Modelling Using Artificial Neural Networks (Anns) Model. *Int.J.Curr.Microbiol.App.Sci.* 9(3) (2020) 127-135.
- [8] R. L. Linsley, Rainfall-Runoff Models An Overview in Rainfall-Runoff Relationship. in: *Proceedings of the International Symposium on Rainfallrunoff Modelling*, May, 18-21 (1982) 3-22.
- [9] M. B. Mcpherson, Some Notes on the Rational Method of Storm Drain Design. Tech. Memo. No. 6 Asce, Water Resources Research Program, Harvard University, Cambridge, Ma. (1969).
- [10] D. R. Maidment, *Handbook of Hydrology*. 1st Edn. New York: Mcgraw Hill Publication. (1993).
- [11] W. H. Green, and G. A. Ampt, Studies on Soil Physics. the Flow of Air and Water Through Soils. *Journal of Agriculture Science*. 4 (1911) 1-24.
- [12] J. E. Nash, Determination of Runoff From Rainfall. *Institute of Civil Engineering*. 10 (1958) 163-184.

- [13] P. S. Kumbhare, and R. A. Rastogi, Determination of Surface Runoff from Himalayan Watershed using Two Parameter Conceptual Model. in: 21st Annual Convocation of Indian Society of Agricultural Engineering, Indian Agriculture Research Institute, New Delhi. 1984
- [14] V. Kumar, and R. A. Rastogi, Determination of Direct Runoff from a Small Agricultural Watershed. *J of Agric. Engg. Isae.* 26(3) (1989) 223-228.
- [15] B. Lahcene, A New Extended-Gamma Family of Distributions: Properties and Applications. *Journal of Applied Mathematics and Computation* , 5(1) (2021) 9-17.
- [16] S. K. Agarwal, and S. L. Kalla, A Generalized Gamma Distribution and Its Application in Reliability. *Commun. Statist. Theory Meth.*, 25(1) (1996) 201-210.
- [17] E. W. Stacy, and G. A. Mihram, Parameter Estimation for A Generalized Gamma Distribution. *Technometrics*, 7(3) (1965), 349-358.
- [18] E. W. Stacy, A Generalization of the Gamma Distribution. *Annals of Mathematical Statistics*, (33) (1962) 1187-1192.
- [19] G. M. Cordeiro, F. Castellares, L. C. Montenegro, and M. De Castro, the Beta Generalized Gamma Distribution. *Statistics (Berlin)*, (46) (2012): 1-13.
- [20] S. Suksaengrakcharoen and W. Bodhisuwan. A New Family of Generalized Gamma Distribution and Its Application. *Journal of Mathematics and Statistics*, 10(2) (2014) 211-220.
- [21] L. Bachioua, on Extended and Reliability General Mixture Gamma Distribution Model. A Dissertation Submitted to the College of Science, University of Baghdad in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy (Ph.D.) of Science in Mathematics, University of Baghdad, Iraq. 2004
- [22] L. Bachioua, and A. M. Mohammed, on a Five and Six Parameter Generalization of the Gamma Function. *Journal Theoretical Mathematics & Applications.* 7(1) (2017) 1-14.
- [23] L. Bachioua, Extended Generalized Gamma Function and Same Its Applications. *International Archive of Applied Sciences and Technology*, 4(3) (2013) 16-30.
- [24] J. Piantadosi, J. Boland, and P. Howlett, Generating Synthetic Rainfall on Various Timescales—Daily, Monthly and Yearly. *Environ Modeling & Assessment*, 14 (2009) 431–438
- [25] D. A. Woolhiser, and J. Roldan, Stochastic Daily Precipitation Models 2. A Comparison of Distribution of Amounts. *Water Resour. Research.* 18(5) (1982) 1461-1468
- [26] D. S. Wilks, Interannual Variability and Extreme Value Characteristics of Several Stochastics Daily Models. *Agricultural and Forest Meteorology.* 93(3) (1999) 153-169
- [27] B. Kedem, L. S. Chiu, and Z. Karni. An Analysis of the Threshold Method for Measuring Area-Average Rainfall. *Journal of Applied Meteorolgy*, 29(1) (1990) 3-20
- [28] R. Shanker, and F. Hagos, On Modeling of Lifetimes Data using Exponential and Lindley Distributions, *Biometrics & Biostatistics International Journal.* 2(5) (2015) 1-9
- [29] R. Shanker, Sujatha Distribution and Its Applications. *Statistics in Transition-New Series.* 17(3) (2016) 1-20
- [30] R. Shanker, Rani Distribution and Its Application. *Biometrics & Biostatistics International Journal.* 6(1) (2017) 256-265.  
Doi: 10.15406/Bbij.2017.06.00155