A new probability model with application to heavy-tailed hydrological data

Tassaddaq Hussain[1] · Hassan S. Bakouch[2] · Christophe Chesneau[3]

Abstract

Because of the dramatic changes that are being observed in the climatic conditions of the world, such as excess of rains, drought and huge floods, we introduce a versatile hydrologic probability model with three parameters. The proposed model is a combination of the Lomax and generalized Weibull distributions based on an exponent odd function. Main properties of the distribution are obtained, such as shapes of the probability density and hazard rate functions, quantile function, asymptotic distribution, information matrix and characterization via hazard rate function. Parameters are estimated via the maximum likelihood estimation method. Four data sets are used to compare the proposed model with a number of well-known hydrologic models. The proposed model is found to be suitable and representative for heavy-tailed hydrological data sets, with least loss of information attitude and a realistic return period.

Keywords Estimation · Goodness-of-fit statistics · Hazard rate function · Statistical distributions · Transformation

Mathematics Subject Classification 60E05 · 62E15 · 62F10

✉ Tassaddaq Hussain
  taskho2000@yahoo.com

  Hassan S. Bakouch
  hassan.bakoush@science.tanta.edu.eg

  Christophe Chesneau
  christophe.chesneau@unicaen.fr

[1] Department of Mathematics, Mirpur University of Science and Technology (MUST), Mirpur 10250, Pakistan

[2] Department of Mathematics, Faculty of Science, Tanta University, Tanta, Egypt

[3] LMNO, University of Caen, Caen, France

⚫ Springer

1 Introduction with the proposed model

Frequency of heavy precipitation or proportion of total rainfall from heavy falls will increase in the 21st century over many areas on the globe, this will increase the likelihood of floods and devastation of infrastructures (Intergovernmental Panel on Climate Change IPCC 2012). These upcoming future devastating effects have certainly opened the hidden corner for accurate modeling of the flood/precipitation/earth quake data which will not only produce good fit but also yield realistic return periods. Such requisition for modeling the hydrologic phenomenon needs to assess the tail behavior which is the only realistic knowledge regarding the behavior of the distribution beyond the range of the sample (Markovich 2007). Such tail behavior has its own significance in its relative discipline. For example, light tail distributions (kurtosis less or equal to 3) can effectively be used in the analysis of global warming in terms of extreme warm and cold temperature (Loikith and Neelin 2015) and rainfall data analysis (Papalexiou et al. 2013), while the heavy-tailed distributions are used to model the service time and input in queuing models, flood levels of rivers, major insurance claims, wave heights during a storm, and low and high temperatures (Markovich 2007). These heavy-tailed distributions include the Pareto, the lognormal, the Weibull with shape parameter less than 1, the Cauchy, the Burr and the Fréchet, while the light-tailed distributions include the exponential, the gamma, the Weibull with shape parameter greater than 1, and the normal distributions.

In this regard, the conventional frequency analysis (CFA) is used for modeling these extreme events via heavy-tailed distributions which include the the Log-Pearson type III (LP(3)) defined by Bobee (1975), the three parameter log-normal distribution (LN(3)) as stated by Krige (1960), the generalized Pareto (GP) distribution used by Hosking and Wallis (1987) and Dargahi-Noubary (1989), the generalized logistic (GLO) distribution studied by Dyrrdal (2012) and Balakrishnan and Leung (1988), Gumbel distribution known as the generalized extreme value type I (GEV-I) distribution investigated by Mujere (2011), the three parameter Kappa distribution studied by Junior and Johnson (1973), the gamma distribution, the generalized exponentiated exponential Lindley (GEEL) proposed by Hussain et al. (2018) and the Fréchet distribution discussed by Ramos et al. (2018).

Moreover, the domain of attractions for maxima–minima, of the LP(3), LN(3), GP, GLO and GEEL distributions belong to the Gumbel–Weibull, the Gumbel–Gumbel, the Fréchet-Weibull, the Gumbel–Gumbel and the Gumbel–Fréchet distributions, respectively. However, the Gumbel distribution is not considered an appropriate model for the analysis of extreme events because its right tail is light. Furthermore, in extreme events analysis, researchers generally prefer the positive skewness coupled with the heavy-tailed phenomena instead of light or intermediate tails distributions (Hussain et al. 2018).

Here, we study a phenomenal procedure designed for modeling heavy-tailed environmental data sets, based on flexible parametric modeling by adding shape, location or scale parameters, while constraining the number of parameters to be at most equal to four. However, three is a recommended number of parameters to draw valid conclusions (Mudholkar et al. 1996, Johnson et al. 2005), which provides a broader range of hazard shapes and allow an assessment of each competing model relative to a more

comprehensive one (Mudholkar et al. 1996). Also, such addition of extra parameter is useful in exploring tail behavior which not only improves the goodness of fit of the proposed models but also provides lower information criteria. Moreover, the exploration of tails usually reflects the importance of extreme and intermediate order statistics or exceedances over high thresholds. Focusing our attention on tails may advantageously inspire the introduction of new parametric statistical models.

The first motivational ground of modeling the extreme events and heavy-tailed phenomenon is based on the selection of Pareto Type II distribution, which was first proposed by Lomax (1987). It is effectively used in reliability modelling, life testing, income, wealth, business failure, firm size, queuing problems, biological sciences, modeling the distribution of the sizes of computer files on servers and as an alternative to the exponential distribution when the data are heavy-tailed data. The Lomax distribution is defined as follows.

Definition 1 A random variable T has the Lomax distribution with shape parameter $\beta > 0$ if its cumulative distribution function (cdf) is given by

$$\pi(t) = 1 - (1+t)^{-\beta}, \quad t > 0, \ \beta > 0,$$

and its corresponding probability density function (pdf) is expressed as

$$\ell(t) = \beta(1+t)^{-\beta-1}, \quad t > 0, \ \beta > 0.$$

The second motivational ground is to handle the shortcomings of Weibull distribution (an asymptotic version of one of the three heavy-tailed phenomena) which is frequently used in statistical analysis of many practical data, however it is inappropriate when the failure rate is indicated to be unimodal and bathtub shaped as well as when data exhibit high kurtosis usually greater (or equal) twenty. The third motivation is the selection of two parameter generalized Weibull (GW); possessing a bathtub hazard rate, simple structure and heavy-tailed behavior, defined by Mudholkar et al. (1996) as a baseline distribution whose cdf defined on the interval $(0, \infty)$ as $G(x) = 1 - \left(1 - \lambda x^\theta\right)^{\frac{1}{\lambda}}$ with the pdf $g(x) = \theta x^{\theta-1} \left(1 - \lambda x^\theta\right)^{\frac{1}{\lambda}-1}$ for $x > 0, \lambda \leq 0$. The fourth motivational aspect is the choice of an appropriate link function, with its help we can overcome the scaling issues of environmental data sets $D(.) : [a, b] \rightarrow [0, \infty]$ which satisfies the next conditions: (i) $D(.)$ is differentiable and monotonically non-decreasing (ii) $D(x) \rightarrow a$ as $x \rightarrow 0$ and $D(x) \rightarrow b$ as $x \rightarrow \infty$, and can be considered as exponent odd function $D(x) = (e^{\theta \frac{G(x)}{1-G(x)}} - 1)/(e^\theta - 1), \theta > 0$; where θ and $e^\theta - 1$ and behave as scale parameters which usually help to overcome the scaling issues of the environmental data sets. Now on the basis of above four motivational aspects, we define the cdf of Lomax D-GW abbreviated as LDGW distribution as follows.

Definition 2 If X is a continuous random variable, then the cdf of the LDGW distribution with parameters λ, θ, β is defined as

$$F_{LDGW}(x|\lambda, \theta, \beta) = \int_0^{D(x)} \ell(t)dt = 1 - \left(1 + \frac{e^{-\theta+\theta\left(1-\lambda x^\theta\right)^{-\frac{1}{\lambda}}} - 1}{e^\theta - 1}\right)^{-\beta}, \quad (1)$$

where $\beta > 0, \theta > 0, \lambda \leq 0$ and $x > 0$.

The corresponding probability density function (pdf) of the LDGW distribution is expressed as

$$f_{LDGW}(x|\lambda, \theta, \beta)$$
$$= \beta\theta^2 \left(1 + \frac{e^{-\theta+\theta\left(1-\lambda x^\theta\right)^{-\frac{1}{\lambda}}} - 1}{e^\theta - 1}\right)^{-\beta-1} \frac{e^{-\theta+\theta\left(1-\lambda x^\theta\right)^{-\frac{1}{\lambda}}} x^{\theta-1} \left(1 - \lambda x^\theta\right)^{-\frac{1}{\lambda}-1}}{e^\theta - 1}.$$
$$(2)$$

Similarly the survival function (sf) of the LDGW distribution is

$$S_{LDGW}(x|\lambda, \theta, \beta) = \left(1 + \frac{e^{-\theta+\theta\left(1-\lambda x^\theta\right)^{-\frac{1}{\lambda}}} - 1}{e^\theta - 1}\right)^{-\beta} \quad (3)$$

and the hazard rate function (hrf) of the LDGW distribution is given by

$$h_{LDGW}(x|\lambda, \theta, \beta) = \frac{\beta\theta^2 e^{-\theta+\theta\left(1-\lambda x^\theta\right)^{-\frac{1}{\lambda}}} x^{-1+\theta} \left(1 - \lambda x^\theta\right)^{-1-\frac{1}{\lambda}}}{e^{-\theta+\theta\left(1-\lambda x^\theta\right)^{-\frac{1}{\lambda}}} + e^\theta - 2}. \quad (4)$$

In addition, we have observed that $\lim_{x\to\infty} e^{\alpha x} S_{LDGW}(x|\lambda, \theta, \beta) = \infty$, for all $\alpha > 0$ and $\theta < -\lambda$. Furthermore, $\lim_{x\to\infty} h_{LDGW}(x|\lambda, \theta, \beta) = 0$ for $\theta < -\lambda$. Hence, the LDGW distribution has a heavy tail when $\theta < -\lambda$ (Foss et al. 2011). This behavior of such distribution adapts hydrological data with heavy tails in the application portion of this paper. Other motivational factors include the handling of floods devastation effects, generally, floods are caused by the heavy concentrated rainfall, which are sometimes augmented by snowmelt flows in rivers. Such floods usually become the causes of financial loss, destruction of infrastructure and spreadness distance issue. In order to cope with the above mentioned causes our proposed model is desired which not only addresses flood and precipitation data at a site or a region but also produces a realistic return period, see real life application of the four data sets as mentioned in the application portion of this paper. Moreover, the proposed model exhibits not only increasing but also upside down bathtub shapes, see Figs. 1 and 2. In addition to, the proposed model exhibits, positive skewness, symmetry and negative skewness coupled with leptokurtic, mesokurtic and plattykurtic behavior (see Figs. 3, 4 and 5). Last but not the least, the asymptomatic distribution of both maxima and minima lies in the Weibull domain of attraction which rarely exists in the literature.

The contents in the rest of this article are arranged as follows. Section 2 deals with some statistical properties of the LDGW distribution, such as quantile function,
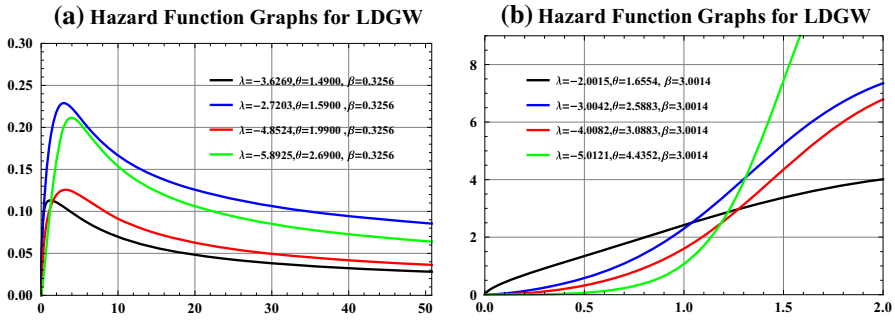
(a) Hazard Function Graphs for LDGW

(b) Hazard Function Graphs for LDGW

Fig. 1 LDGW hrf graphs for the indicated values

(a) Hazard Function Plot for LDGW

(b) Hazard Function Graphs for LDGW

Fig. 2 LDGW hrf graphs for the indicated values

(a) Probability Function Graphs for LDGW

(b) Probability Function Graphs for LDGW
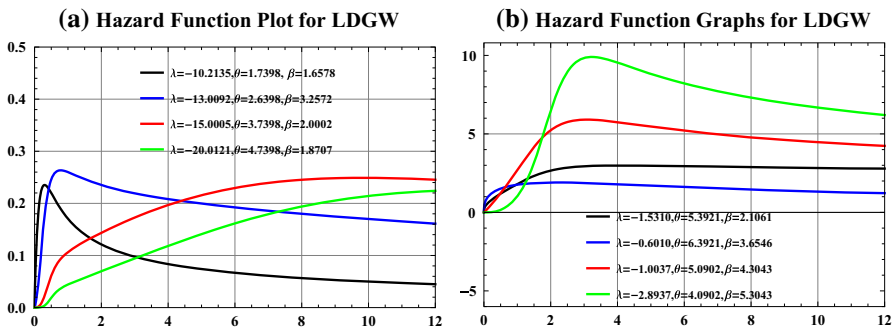
Fig. 3 LDGW pdf graphs for the indicated values

asymptotic distributions of largest and smallest order statistics and characterization issues via the hazard function. Parameters estimation by maximum likelihood is studied in Sect. 3 as well as hydrological applications are explored and comparison is made with reputed hydrological probability models. Concluding remarks, findings and recommendation are given in Sect. 4.

Fig. 4 LDGW pdf graphs for the indicated values

2 Some properties of the model

Here, we shall discuss some properties of the LDGW distribution.

2.1 Quantile function

If X is an absolutely continuous random variable having the pdf defined in (2), then its quantile function, say Q_p (qu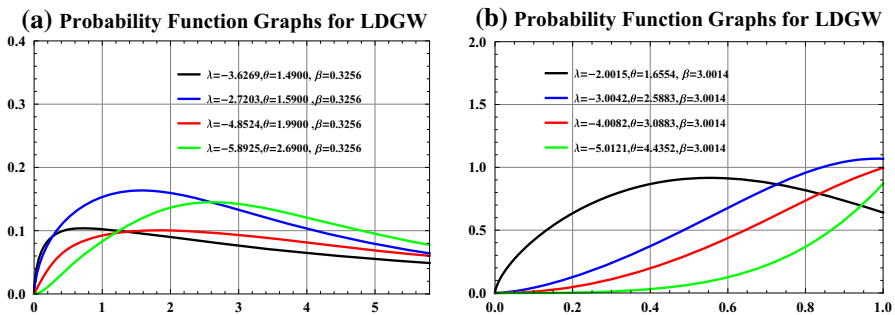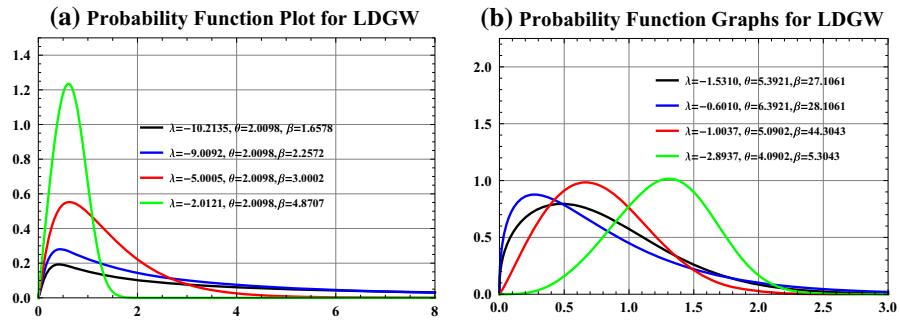antile of order p), is defined as $F_{LDGW}(Q_p | \lambda, \theta, \beta) = p$, where $0 < p < 1$. Quantile function gives a number Q_p in such a way that area to the left of Q_p is p, i.e., it is the root of the function

$$Q_p = \lambda^{-\frac{1}{\theta}} \left(1 - \left(\frac{\theta + \ln \left\{ 2 - e^{\theta} - (1 - e^{\theta})(1 - p)^{-\frac{1}{\beta}} \right\}}{\theta} \right)^{-\lambda} \right)^{\frac{1}{\theta}}. \tag{5}$$

Equation (5) can be used to simulate LDGW random variables.

Now by using (5) we can find the median, skewness and kurtosis of LDGW as Median $= Q_{0.5}$, when $p = 0.5$,

$$Sk. = \frac{Q_{\frac{3}{4}} - 2Q_{\frac{1}{2}} + Q_{\frac{1}{4}}}{Q_{\frac{3}{4}} - Q_{\frac{1}{4}}}, \tag{6}$$

$$Ku. = \frac{Q_{\frac{7}{8}} - Q_{\frac{5}{8}} + Q_{\frac{3}{8}} - Q_{\frac{1}{8}}}{Q_{\frac{6}{8}} - Q_{\frac{2}{8}}}. \tag{7}$$

Equation (7) indicates that: As kurtosis increases, the tail of the distribution becomes heavier. These measures are less sensitive to outliers and they exist even for distributions without moments (Cakmakyapan and Ozel 2016). Figure 5 gives the plots of skewness and kurtosis for the LDGW distribution. From this figure, the distribution is leptokurtic and mesokurtic as well as platykurtic attitudes. From this, it is evident the
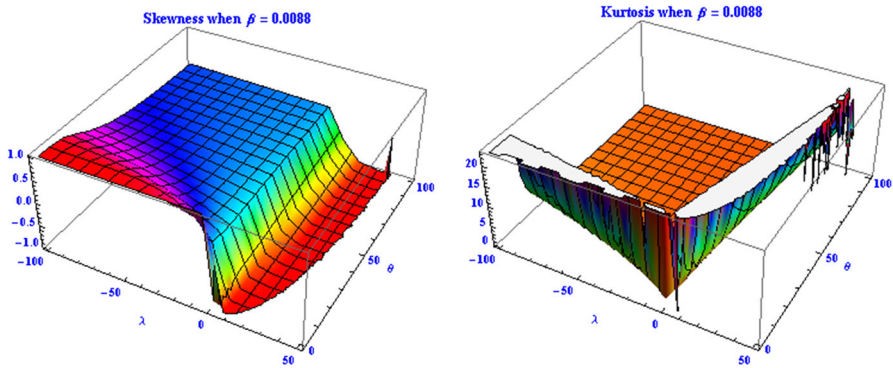
Fig. 5 Skewness and Kurtosis of the LDGW distribution for the indicated values

distribution reflects light and heavy tail behavior. Moreover, it is also observed that kurtosis and skewness decrease as β increases.

2.2 Asymptotic distribution

Here, we present the asymptotic distributions of $W_n = X_{n:n}$ and $w_n = X_{1:n}$ that are the largest and smallest observations, respectively, from the random sample of n observations. For this purpose we take the cdf and pdf of the LDGW distribution. Then, when $\theta < -\lambda$, the distribution of maxima is

$$\lim_{t \to \infty} \frac{1 - F_{LDGW}(t + x|\lambda, \theta, \beta)}{1 - F_{LDGW}(t|\lambda, \theta, \beta)} = \lim_{t \to \infty} \frac{S_{LDGW}(t + x|\lambda, \theta, \beta)}{S_{LDGW}(t|\lambda, \theta, \beta)} = 1.$$

Hence, the tail of $F_{LDGW}(.)$ is of slow variation. Also, the distributions with slowly varying tails are shown to be of valuable use in practice (Alves et al. 2009). So, the standardized form of distribution of maxima is Weibull (type -III). Now, the distribution of minima for all values of λ, after using L'Hôpital's rule, is

$$\lim_{t \to 0} \frac{F_{LDGW}(tx|\lambda, \theta, \beta)}{F_{LDGW}(t|\lambda, \theta, \beta)} = \lim_{t \to 0} \frac{x f_{LDGW}(tx|\lambda, \theta, \beta)}{f_{LDGW}(t|\lambda, \theta, \beta)} = x^{\theta},$$

and its standardized form is Weibull (type -III), i.e., $1 - e^{-x^{\theta}}$. According to (Leadbetter et al. 1987, Theorem 1.6.2) of norming constants $a_n > 0$, $b_n > 0$, $c_n > 0$ and $d_n > 0$, we obtain

$$P\{a_n(W_n - b_n) \le x\} \to e^{-1}, \quad P\{c_n(w_n - d_n) \le x\} \to 1 - e^{-x^{\theta}},$$

as $n \to \infty$. From this, it is clear that the asymptotic distribution of sample maxima and sample minima is in the domain of attraction of Weibull distribution.

2.3 Characterization based on the hazard rate function

Characterization usually helps the researchers for determining the exact probability distribution. The hrf is twice differentiable and usually satisfies the first order differential equation

$$\frac{f'(x)}{f(x)} = \frac{h'_f(x)}{h_f(x)} - h_f(x),$$

where $h_f(x) = \dfrac{f(x)}{1 - F(x)}$ is the hrf. The following characterization establishes a non-trivial characterization for the LDGW distribution in terms of the hazard rate function.

Theorem 1 *Let $X : \Omega \to (0, \infty)$ be an absolutely continuous random variable possessing sf $S_F(x)$ and $f(x)$. Then it is said to have pdf as defined in Eq. (2) iff it satisfies the expression*

$$\frac{(\ln h(x))'}{h(x)} = \left\{ \theta^2 x^{\theta-1}(1 - \lambda x^\theta)^{-\frac{1}{\lambda}-1} + \frac{\theta - 1}{x} + \frac{\lambda \theta x^{\theta-1}(1 - \lambda x^\theta)^{-1}(1 + \lambda)}{\lambda} \right.$$

$$\left. - \frac{(\beta + 1)\theta^2 x^{\theta-1}(1 - \lambda x^\theta)^{-\frac{1}{\lambda}-1} e^{-\theta + \theta(1-\lambda x^\theta)^{-\frac{1}{\lambda}}}}{e^{-\theta + \theta(1-\lambda x^\theta)^{-\frac{1}{\lambda}}} + e^\theta - 2} \right\} h(x) + 1, \qquad (8)$$

for $\beta > 0$, $\theta > 0$, $\lambda \leq 0$ and $x > 0$.

The proof is postponed in "Appendix-A".

3 Maximum likelihood estimation and hydrological applications

3.1 Maximum likelihood estimation

Let $X_1, X_2, \ldots, X_n$ be a random sample from the distribution characterized by (2) with parameter vector $\mathbf{\Theta} = (\lambda, \beta, \theta)$ and $x_1, x_2, \ldots, x_n$ are the corresponding observed values, then its log-likelihood is expressed as

$$\ell(\Theta) = \ln[L(x_1, x_2, \ldots x_n | \Theta)] = (\theta - 1) \sum_{i=1}^{n} \ln(x_i) - n \ln(e^\theta - 1)$$

$$- n\theta + \theta \sum_{i=1}^{n} \left(1 - \lambda x_i^\theta\right)^{-\frac{1}{\lambda}}$$

$$+ n \ln(\beta) + 2n \ln(\theta) - \left(1 + \frac{1}{\lambda}\right) \sum_{i=1}^{n} \ln(1 - \lambda x_i^{\theta})$$

$$- (\beta + 1) \sum_{i=1}^{n} \ln \left[1 + \frac{-1 + e^{-\theta + \theta\left(1 - \lambda x_i^{\theta}\right)^{-\frac{1}{\lambda}}}}{e^{\theta} - 1} \right]. \tag{9}$$

The maximum-likelihood estimators (MLEs) of λ, θ and β are obtained by solving numerically the nonlinear system of equations $\frac{\partial \ell(\Theta)}{\partial \theta} = 0$, $\frac{\partial \ell(\Theta)}{\partial \lambda} = 0$ and $\frac{\partial \ell(\Theta)}{\partial \beta} = 0$. Those partial derivatives are given in "Appendix-B". Moreover, for determining the variance co-variance matrix and the confidence interval for the distribution parameters, one requires information matrix which can be generated by the taking the expectation of the second order derivative. In this regard the second order derivatives of the equations above can be provided on demand.

3.2 Evaluation tests

In order to demonstrate the proposed methodology, we consider four different real-world data sets, and compare ten distributions which are too much popular in hydrologic data analysis including, the Weibull, generalized Pareto (3), log-normal (3), log Pearson Type 3, Kappa(3), extreme value, Fréchet, generalized logistic generalized exponentiated exponential Lindley (GEEL) and generalized Weibull distributions, for probability density functions of these distributions readers referred to Hussain et al. (2018). The statistics: Akaike information criterion (AIC), corrected Akaike information criterion (AICc), Hannan–Quinn information criterion (HQIC) and consistent Akaike information criterion (CAIC) along with Vuong test are used to select the best model among several models. The definitions of AIC, AICc, HQIC and CAIC are given as:

$$AIC = 2k - 2\ell(\hat{\Theta}), \qquad AICc = AIC + \frac{2k(k+1)}{n-k-1},$$

$$HQIC = -2\ell(\hat{\Theta}) + 2k \ln(\ln(n)), \qquad CAIC = -2\ell(\hat{\Theta}) + \frac{2kn}{n-k-1}.$$

Moreover, perfection of competing models is also tested via the Kolmogrov−Simnorov(K-S), the Anderson–Darling (A_0^*) and the Cramer Von Misses (W_0^*) statistics. The mathematical expressions for the statistics above are given by

$$KS = \max \left\{ \frac{i}{m} - z_i, z_i - \frac{i-1}{m} \right\}, \chi^2 = \sum_{i=1}^{m} \frac{(o_i - e_i)^2}{e_i},$$

$$A_0^* = \left(\frac{2.25}{m^2} + \frac{0.75}{m} + 1 \right) \left\{ -m - \frac{1}{m} \sum_{i=1}^{m} (2i-1) \ln(z_i (1 - z_{m-i+1})) \right\},$$

$$W_0^* = \sum_{i=1}^{m} \left(z_i - \frac{2i-1}{2m} \right)^2 + \frac{1}{12m},$$

where k denotes the number of parameters, n denotes the number of observations, m denotes the number of classes, $z_i = \mathfrak{Q}(x_i)$, the x_i's are the ordered observations, o_i and e_i are the observed and expected frequencies of the ith class, respectively. Whereas the Vuong test with its concerning procedure is outlined as follows.

Vuong test Chi-square approximation in the regard of the likelihood ratio test statistic is valid only for testing restrictions on the parameters of a statistical model (i.e., $\mathbf{H}_0$ and $\mathbf{H}_1$ are nested hypotheses). However when models are non-nested, we can not use the likelihood ratio tests for model comparison. In this scenario, the AIC AICc, CAIC and HQIC as well as the Vuong test for non-nested models are useful. Vuong (1989) proposed a likelihood ratio-based statistic for testing the null hypothesis that the competing models are equally close to the true data generating process against the alternative that one model is closer. Let us consider two statistical models based on the probability density functions $f_A(x; \xi) = F_A(b) - F_A(a)$ and $f_B(x; \varrho) = F_B(b) - F_B(a)$ possess equal or unequal number of parameters. Now, we define the likelihood ratio statistic for the model $f_1(x; \xi)$ against $f_2(x; \varrho)$ as

$$\mathfrak{LR}(\hat{\xi}_n, \hat{\varrho}_n) = \sum_{k=0}^{k_{max}} f_k ln\left(\frac{f_A(x; \hat{\xi}_n)}{f_B(x; \hat{\varrho}_n)}\right),$$

where $\hat{\xi}_n$ and $\hat{\varrho}_n$ are the MLEs in each model based on the sample $k_1, k_2, \ldots k_n$ and f_k denotes the observed frequency (Denuit et al. 2007). If both models are strictly non-nested, then under $\mathbf{H}_0$ we have the test statistic $\mathbf{Z} = \frac{\mathfrak{LR}(\hat{\xi}_n, \hat{\varrho}_n)}{\sqrt{n}\hat{\omega}_n}$, having the $\mathcal{N}(0, 1)$ distribution as approximated distribution when n is large, where

$$\hat{\omega}_n = \frac{1}{n}\sum_{i=1}^{n}\left(ln\frac{f_A(x_i; \hat{\xi}_n)}{f_B(x_i; \hat{\varrho}_n)}\right)^2 - \left(\frac{1}{n}\sum_{i=1}^{n}ln\frac{f_A(x_i; \hat{\xi}_n)}{f_B(x_i; \hat{\varrho}_n)}\right)^2.$$

In order to select an appropriate model we usually construct the following hypothesis

$$\mathbf{H}_0 : \mathfrak{KL}(f_A(x; \xi)) = \mathfrak{KL}(f_B(x; \varrho)),$$

that is, the model A is defined to be better than model B if model A's $\mathfrak{KL}$ distance to the truth is smaller than the model B,

$$\mathbf{H}_{1A} : \mathfrak{KL}(f_A(x; \xi)) < \mathfrak{KL}(f_B(x; \varrho)),$$

with corresponding critical value, i.e., we reject $\mathbf{H}_0$ when $\mathbf{Z} > \mathbf{Z}_\gamma$ in favor of A. Similarly, model B is defined to be better than model A if model B's $\mathfrak{KL}$ distance to the truth is smaller than the model A,

$$\mathbf{H}_{1B} : \mathfrak{KL}(f_A(x; \xi)) > \mathfrak{KL}(f_B(x; \varrho)),$$

with the corresponding critical region, i.e., we reject $\mathbf{H}_0$ when $\mathbf{Z} < -\mathbf{Z}_\gamma$ in favor of B otherwise decision can not be made, where $\mathfrak{KL}$ denotes the Kullback–Leibler

Table 1 Descriptive statistics for data set I

Sample size	Median	Skewness	Kurtosis	$\frac{Skewness}{Kurtosis}$	Sk_B	Ku_B
39	3570	4.5581	25.4436	0.1791	0.2457	1.7548

Table 2 Some theoretical measures for data set I from LDGW

Sample size	Median	Skewness	Kurtosis	$\frac{Skewness}{Kurtosis}$	Sk_B	Ku_B
39	3306.69	2.1644	8.7915	0.2461	0.3688	1.7926

distance measure and γ denotes the level of significance. Therefore, if the value of the test statistic is higher than $\mathbf{Z}_\gamma$ then one can reject the null hypothesis that the models are equivalent in favor of $f_A(x; \xi)$ being better than $f_B(x; \varrho)$, and if the test statistic is smaller than $-\mathbf{Z}_\gamma$ then one rejects the null hypothesis in favor of $f_B(x; \varrho)$ being better than $f_A(x; \xi)$ else decision can not be made. The comparisons of the competing models are displayed in the Table 17 for the four data sets.

Example 1 The first data set (I) is the annual flood discharge rates of the Floyd River data flood discharge (ft^3/s) during 1935–1973, and it was first reported by Pickands (1975) and later on by Akinsete et al. (2008). The data consist of the following values: 1460, 4050, 3570, 2060, 1300, 1390, 1720, 6280, 1360, 7440, 5320, 1400, 3240, 2710, 4520, 4840, 8320, 13900, 71500, 6250, 2260, 318, 1330, 970, 1920, 15100, 2870, 20600, 3810, 726, 7500, 7170, 2000, 829, 17300, 4740, 13400, 2940, 5660.

Analysis of data set I: It is evident from Table 1 that the data behave as positively skewed with high kurtosis in such a way that skewness to kurtosis ratio is 0.1791. Furthermore, Bowley skewness to Bowley kurtosis ratio based on quantiles is also worth noting which is 0.1400. So, such description of the data demands a model which can work very well in positively skewed and leptokurtic distributions. Moreover, Table 2 also affirms the above statement. To draw a valid conclusion we have converted the ungrouped data into grouped data by using the command **b**ins of **R** computational package (Tables 3 and 4). For this purpose we have created different classes, such as, $[318, 1.37 \times 10^3]$, $(1.37 \times 10^3, 2.04 \times 10^3]$, $(2.04 \times 10^3, 3.57 \times 10^3]$, $(5.43 \times 10^3, 8.05 \times 10^3]$, $(8.05 \times 10^3, 7.15 \times 10^4]$, $(7.15 \times 10^4, \infty)$ with observed frequencies of each class, which are 7, 6, 7, 6, 6, 7 respectively. So in this regard Tables 3 and 4 portray the comparison of the compared distributions which are highly recommended in heavy-tailed behavior. From this table, it is evident that our proposed model is the most suitable one, with least values for all statistics and highest p value for χ^2 statistics. Thus showing that there is a close association between the data and the proposed model.

Example 2 The second data set (II) is the Wabash River Flows at Mt. Carmel (Threshold = 50,000 cfs). The data cover a period of 65 years (1928–1992) with 281 peaks that exceed 50,000 cfs with an average annual number of peaks $281/65 = 4.32308$. This data set was reported by (Rao and Hameed 2000, page 297).

Table 3 Maximum likelihood estimators and goodness of fit statistics of data set I

Distribution	$\hat{\lambda}$	$\hat{\beta}$	$\hat{\theta}$	χ^2(d.f)	p value	KS	A_0^*	W_0^*
Weibull(β,θ)	–	0.8715	6189.6259	3.5429(3)	0.315	0.1271	1.0754	0.1513
Fr.(β,θ)	–	1.0145	2151.5901	1.7353(4)	0.784	0.0862	0.3938	0.0502
EV(β,θ)	–	3552.7109	4246.2181	12.4091(3)	0.006	0.1631	2.3600	0.3886
GLO(β,θ)	–	3.0876	0.0003	16.9198(3)	0.001	0.1808	2.8111	0.4705
GP(λ,β,θ)	2600.0656	0.7125	0.5209	0.9642(3)	0.810	0.0724	0.1935	0.0290
GW(λ,β,θ)	-2.7775	4.7693×10^{-8}	2.3073	8.2351(4)	0.083	0.2263	3.3010	0.6698
LN3(λ,β,θ)	8.1222	1.1198	127.9900	0.7133(3)	0.870	0.0861	0.3924	0.0500
GEEL(λ,β,θ)	5.0453	1.8752	0.0003	10.6837(3)	0.014	0.2457	6.0615	0.9768
Kappa3(λ,β,θ)	0.00096	2152.9316	1063.1286	1.7294(4)	0.785	0.0861	0.3924	0.0501
LP3(λ,β,θ)	4.9874	0.7621	4.8649	5.8554(3)	0.119	0.2077	2.2596	0.3976
LDGW(λ,β,θ)	-144.6139	0.5006	12.5202	1.0421(4)	0.903	0.0722	0.1985	0.0308

Table 4 Log likelihood (l)and information criterion for data set I

Distribution	$-l$	AIC	AICC	HQIC	CAIC
Weibull(β,θ)	382.1299	768.2598	765.5675	769.4535	768.5931
Fr.(β,θ)	377.9868	759.9736	757.2813	761.1673	760.3069
EV(β,θ)	394.3573	792.7146	790.0223	793.9083	793.0479
GLO(β,θ)	398.1236	800.2472	797.5549	801.4409	800.5805
GP(λ,β,θ)	376.6006	759.2012	755.8166	760.9918	759.8869
GW(λ,β,θ)	381.4146	768.8292	765.4446	770.6198	769.5149
LN3(λ,β,θ)	376.5220	759.044	755.6594	760.8346	759.7297
Kappa(λ,β,θ)	377.9787	761.9574	758.5728	763.748	762.6431
GEEL(λ,β,θ)	392.6568	791.3136	787.929	793.1042	791.9993
LP3(λ,β,θ)	381.7355	769.471	766.0864	771.2616	770.1567
LDGW(λ,β,θ)	376.6024	759.2048	755.8202	760.9954	759.8905

Analysis of data set II: As Table 5 portrays that the data are positively skewed with high kurtosis. We have fitted all of the probability distributions by using the MLEs. The descriptive statistics are as displayed in Table 5 and the corresponding theoretical ones in Table 6.The skewness to kurtosis ratio is 0.1452. Among the competitive models the proposed model exhibits smaller values for all of the above defined statistics. For obtaining the χ^2 statistics we have created 10 classes via **R** computational package (R Core Team 2013). The created classes are $[1.26 \times 10^4, 5.48 \times 10^4,(5.48 \times 10^4, 6.07 \times 10^4],(6.07 \times 10^4, 6.98 \times 10^4],(6.98 \times 10^4, 7.87 \times 10^4],(7.87 \times 10^4, 8.67 \times 10^4],(8.67 \times 10^4, 1 \times 10^5],(1 \times 10^5, 1.14 \times 10^5],(1.14 \times 10^4, 1.34 \times 10^5],(1.34 \times 10^5, 1.60 \times 10^5],(1.60 \times 10^5, 5.50 \times 10^5]$ while observed frequencies are 30, 27, 28, 28, 28, 28, 28, 31, 25, 28, respectively. The calculated χ^2 for the proposed model is the least with highest p value thus indicating that the proposed model is the suitable one for such data set (Tables 7 and 8).

Table 5 Descriptive statistics for wabash river flows at Mt. Carmel

Sample size	Median	Skewness	Kurtosis	$\frac{Skewness}{Kurtosis}$	$Sk_{B.}$	$Ku_{B.}$
281	86700	2.9551	20.3424	0.14526	0.2749	1.0943

Table 6 Some theoretical measures for wabash river flows at Mt. Carmel from LDGW

Sample size	Median	Skewness	Kurtosis	$\frac{Skewness}{Kurtosis}$	$Sk_{B.}$	$Ku_{B.}$
281	87555.6	2.2956	10.274	0.2234	0.2037	1.4592

Furthermore, information criteria also exhibit least lost of information attitude. This indicates that proposed model use each value of the data set in an efficient manner which is not observed in other competing models.

Example 3 The third data set (III) is the maximum amount of rain in mm of Pakistani city Kalat. The data covers a period of 30 years (1981–2010) with 30 values of maximum rain fall, it was reported by Ahmad et al. (1988). The data are as follows: 32.77, 58.65, 60.71, 64.01, 42.6, 75.8, 88.6, 90.1, 97.9, 105.6, 73.1, 76.6, 78.5, 58.3, 122.5, 57.8, 546, 125.8, 50.5, 45.9, 21.7, 45.5, 38, 75.4, 168.2, 72.9, 95.8, 133.4, 71.9, 28.

Analysis of data set III: Tables 9 and 10 depict that the data consist of 30 observations of rain fall data set of Pakistani city Kalat, indicating a positive skewness coupled with high kurtosis. All of the competing probability distributions are fitted by using the MLEs. The descriptive statistics are displayed in Table 9 and the corresponding theoretical ones in Table 10.The skewness to kurtosis ratio is 0.1969. The comparison as given in Table 11 indicates that the proposed model as good as the Kappa(3). Moreover, we calculate the the χ^2 statistics by creating 6 classes via **R** computational package (R Core Team 2013). The created classes are (21.7, 45], (45, 58.65] ,(58.65, 73], (73, 81.9],(81.9, 108], (108, 546] while observed frequencies are 5, 5, 5, 5, 5, 5, respectively.
In addition to, the information criteria also suggest the proposed model exhibits the least loss of information behavior. Moreover, the Vuong test indicates that the proposed model and Dagum as well as Kappa distributions are the strong candidates for this data set. However, information criteria indicate that proposed model uses each value of the data set in an efficient manner which is not observed in other competing models (Table 12).

Example 4 The fourth data set (IV) is the annual maximum flow (in m^3/s) recorded at Kinrara, Spey for the period 1952–1982 and the number of records is 31. The data were reported by Ahmad et al. (1988). The data are as follows: 89.8, 109.1, 202.2, 146.3, 212.3, 116.7, 109.1, 80.7, 127.4, 138.8, 283.5, 85.6, 105.5, 118.0, 387.8, 80.7, 165.7, 111.6, 134.4, 131.5, 102.0, 104.3, 242.5, 214.8, 144.6, 114.2, 98.3, 102.8, 104.3, 196.2, 143.7.

Analysis of data set IV: The descriptive statistics as displayed in Table 13 and the corresponding theoretical ones in Table 14. The skewness to kurtosis ratio is 0.2256,

Table 7 Maximum likelihood estimators and goodness of fit statistics for data set II

Distribution	$\hat{\lambda}$	$\hat{\beta}$	$\hat{\theta}$	χ^2	p value	KS	A_0^*	W_0^*
Weibull(β, θ)	–	1.9958	114996.9017	55.8904(9)	0.000	0.1657	7.0170	1.1341
Fr.(β, θ)	–	1.8651	72998.3199	39.4035(9)	0.000	0.1263	6.5037	0.8223
EV(β, θ)	–	80533.2581	33965.8818	19.2312(9)	0.023	0.1048	3.7615	0.6109
GLO(β, θ)	–	12.7180	0.00003	22.0123(9)	0.009	0.1082	4.2353	0.6759
GP(λ, β, θ)	2.3296×10^8	2615.3540	12,600	164.7626(9)	0.000	0.3362	NA	4.6921
GW(λ, β, θ)	−1267.6563	0.0054	111.0020	4756.4748(2)	0.000	0.6035	89.0300	19.4443
LN3(λ, β, θ)	11.4885	0.4234	−5609.3751	14.9578(9)	0.092	0.0928	2.7786	0.4421
GEEL(λ, β, θ)	9.3019	-4.227×10^{-06}	0.000024	244.0477(7)	0.000	0.3066	63.1852	8.0852
Kappa3(λ, β, θ)	0.00034	62.8798	564.4511	4069.2913(2)	0.000	0.7462	147.7214	31.7273
LP3(λ, β, θ)	11.8607	0.4393	6.8367	332.4933(9)	0.000	0.3701	48.4219	10.0088
LDGW(λ, β, θ)	−690.8240	0.5353	42.1998	13.8995(9)	0.126	0.0651	1.3059	0.1656

Table 8 Log likelihood (l) and information criterion for data set II

Distribution	$-l$	AIC	AICC	HQIC	CAIC
Weibull(β, θ)	3425.4302	6854.86	6851.903	6857.779	6854.904
Fr.(β, θ)	3433.6599	6871.32	6868.363	6874.238	6871.363
EV(β, θ)	3387.7110	6779.422	6776.465	6782.34	6779.465
GLO(β, θ)	3389.8747	6783.749	6780.792	6786.668	6783.793
GP(λ, β, θ)	3483.6953	6973.391	6969.476	6977.768	6973.477
GW(λ, β, θ)	4175.7534	8357.507	8353.592	8361.884	8357.593
LN3(λ, β, θ)	3385.4767	6776.953	6773.039	6781.331	6777.04
GEEL(λ, β, θ)	3392.3002	6790.6	6786.686	6794.978	6790.687
Kappa3(λ, β, θ)	4136.1013	8278.203	8274.288	8282.58	8278.289
LP3(λ, β, θ)	3591.8838	7189.768	7185.853	7194.145	7189.854
LDGW(λ, β, θ)	3380.1537	6766.3070	6762.3930	6770.6850	6766.394

Table 9 Descriptive statistics for maximum amount of rain in mm of Pakistani city Kalat

Sample Size	Median	Skewness	Kurtosis	$\frac{Skewness}{Kurtosis}$	Sk_B	Ku_B
30	73	4.19394	21.2994	0.1969	0.0066	1.5874

Table 10 Some theoretical measures for maximum amount of rain in mm of Pakistani city Kalat from LDGW

Sample Size	Median	Skewness	Kurtosis	$\frac{Skewness}{Kurtosis}$	Sk_B	Ku_B
30	69.9091	1.9674	7.5536	0.2605	0.2179	1.4709

Table 11 Maximum likelihood estimators and goodness of fit statistics for data set III

Distribution	$\hat{\lambda}$	$\hat{\beta}$	$\hat{\theta}$	χ^2	p value	KS	A_0^*	W_0^*
Weibull(β, θ)	–	1.3138	99.3311	3.2853(3)	0.350	0.1293	0.8479	0.1222
Fr.(β, θ)	–	1.8990	54.5467	3.8209(2)	0.148	0.1416	0.6254	0.1057
EV(β, θ)	–	63.7460	37.0421	3.1491(2)	0.207	0.1557	0.8943	0.1217
GLO(β, θ)	–	7.3043	0.0295	3.0876(3)	0.378	0.1615	0.9432	0.1273
GP(λ, β, θ)	451.3255	7.6720	21.7	8.4027(3)	0.038	0.2433	N.A.	0.2433
GW(λ, β, θ)	-1.5226	4.5742×10^{-7}	3.5098	1.3627(4)	0.851	0.1120	0.3116	0.0537
LN3(λ, β, θ)	4.0483	0.7298	12.0047	2.8635(4)	0.581	0.1126	0.4752	0.0713
GEEL(λ, β, θ)	3.5982	-0.0008	0.0226	3.4764(3)	0.324	0.1521	0.8850	0.1305
Kappa(λ, β, θ)	0.6809	66.7702	4.2337	1.0250(4)	0.921	0.0920	0.2254	0.0326
LP3(λ, β, θ)	5.9556	0.2816	2.4852	7.1695(3)	0.067	0.2257	1.6432	0.3413
LDGW(λ, β, θ)	-79.7834	0.5333	12.7016	0.7734(4)	0.942	0.0894	0.2301	0.0323

NA not available

Table 12 Log likelihood (l)and information criterion for data set III

Distribution	$-l$	AIC	AICC	HQIC	CAIC
Weibull(β, θ)	162.8018	329.6036	327.0036	330.5001	330.048
Fr.(β, θ)	154.8099	313.6198	311.0198	314.5163	314.0642
EV(β, θ)	159.6930	323.386	320.786	324.2825	323.8304
GLO(β, θ)	160.6316	325.2632	322.6632	326.1597	325.7076
GP(λ, β, θ)	156.1486	318.2972	315.0972	319.642	319.2203
GW(λ, β, θ)	153.4683	312.9366	309.7366	314.2814	313.8597
LN3(λ, β, θ)	154.5718	315.1436	311.9436	316.4884	316.0667
GEEL(λ, β, θ)	158.1517	322.3034	319.1034	323.6482	323.2265
Kappa3(λ, β, θ)	153.3992	312.7984	309.5984	314.1432	313.7215
LP3(λ, β, θ)	155.8643	317.7286	314.5286	319.0734	318.6517
LDGW(λ, β, θ)	153.4493	312.8987	309.6987	314.2435	313.8218

Table 13 Descriptive statistics for annual maximum flow (in m^3/s) recorded at Kinrara, Spey

Sample Size	Median	Skewness	Kurtosis	$\frac{Skewness}{Kurtosis}$	Sk_B	Ku_B
31	118	1.9129	6.8286	0.2801	0.5155	1.5521

Table 14 Some theoretical measures for annual maximum flow (in m^3/s) recorded at Kinrara, Spey from LDGW

Sample Size	Median	Skewness	Kurtosis	$\frac{Skewness}{Kurtosis}$	Sk_B	Ku_B
31	122.706	2.1423	8.5031	0.2519	0.3072	1.5307

also the data having positive skewness coupled with high kurtosis. The comparison as given in Tables 15 and 16 indicates that the proposed model is the most recommended model with minimum χ^2 and highest p value. The χ^2 statistic is compiled by creating 5 classes via **R** computational package (R Core Team 2013). The created classes (80.7, 102], (102, 109], (109, 118], (118, 144], (144, 202], (202, 388] along with the observed frequencies are 6, 6, 4, 5, 5, 5 respectively.

The proposed model gives least loss of information criteria by depicting minimum values of AIC, AICC, HQIC and AICc. Moreover, the χ^2 indicates that the proposed model is the best one, also it can be seen again this result by looking up for the information criteria of this model.

Comparison via Vuong Test The Vuong test summary is portrayed in Table 17. In this table we have compared the competing distributions with the proposed model at 5 percent level of significance, i.e., Reject H_o if $Z_{Calculated} \geq Z_{0.05} = 1.645$ in favor of model A or Reject H_o if $Z_{Calculated} \leq Z_{0.05} = -1.645$ in favor of model B else decision cannot be made. This table indicates that the proposed model is also a reasonable choice for such data sets by portraying higher Vuong test statistic values, however it also indicates that the GP and LN3 for data set-I and the Kappa3 for data

Table 15 Maximum likelihood estimators and goodness of fit statistics for data set IV

Distribution	$\hat{\lambda}$	$\hat{\beta}$	$\hat{\theta}$	χ^2	p value	KS	A_0^*	W_0^*
Weibull(β, θ)	–	2.2928	164.4540	17.2829(2)	0.000	0.1797	1.4180	0.2191
Fr.(β, θ)	–	3.5883	113.5558	1.4751(2)	0.478	0.1188	0.3488	0.0448
EV(β, θ)	–	118.9713	39.1103	5.6178(2)	0.060	0.1203	0.7395	0.0825
GLO(β, θ)	–	23.4585	0.0263	5.7926(2)	0.055	0.1246	0.7889	0.0853
GP(λ, β, θ)	2370.3742	37.6897	80.7	2.8323(2)	0.243	0.1433	N.A.	0.0774
GW(λ, β, θ)	-2.9999	1.0292×10^{-19}	9.2569	1.6238(2)	0.444	0.1254	0.4514	0.0631
LN3(λ, β, θ)	4.8698	0.3836	3.7702	5.8226(2)	0.054	0.1218	0.6379	0.0832
GEEL(λ, β, θ)	5.4665	88.8688	0.0137	20.3762(2)	0.0001	0.3049	3.0805	0.5670
Kappa(λ, β, θ)	0.3546	119.9933	11.8503	2.6723(2)	0.263	0.1335	0.5093	0.0552
LP3(λ, β, θ)	6.2198	0.1816	3.7702	3.0407(2)	0.219	0.1579	0.7877	0.1137
LDGW(λ, β, θ)	-292.4228	0.1902	43.4412	0.7683(2)	0.681	0.1050	0.3271	0.0524

NA not available

Table 16 Log likelihood (l) and information criterion for data set IV

Distribution	$-l$	AIC	AICC	HQIC	CAIC
Weibull(β, θ)	171.5685	347.137	344.5241	348.0719	347.5656
Fr.(β, θ)	162.1036	328.2072	325.5943	329.1421	328.6358
EV(β, θ)	165.5294	335.0588	332.4459	335.9937	335.4874
GLO(β, θ)	165.8120	335.624	333.0111	336.5589	336.0526
GP(λ, β, θ)	161.5663	329.1326	325.9068	330.5349	330.0215
GW(λ, β, θ)	162.2268	330.4536	327.2278	331.8559	331.3425
LN3(λ, β, θ)	165.2506	336.5012	333.2754	337.9035	337.3901
GEEL(λ, β, θ)	175.8027	357.6054	354.3796	359.0077	358.4943
Kappa3(λ, β, θ)	163.0665	332.133	328.9072	333.5353	333.0219
LP3(λ, β, θ)	164.4232	334.8464	331.6206	336.2487	335.7353
LDGW(λ, β, θ)	161.7888	329.5776	326.3518	330.9799	330.4665

set-III are also reasonable competitor. However, an encouraging aspect of the proposed model is having least loss of information statistics and highest p value which makes it reasonable competitor of the existing models.

For the four data sets, we get the elements of the variance-covariance matrices of the MLES of the LDGW distribution, see Table 18. Using this table, we get the confidence intervals of the parameters of the LDGW distribution for all the data sets, see Tables 19, 20, 21 and 22. Tables 18, 19, 20, 21 and 22 are given in the "Appendix C".

3.3 Hydrological parameters

After observing the suitability of the proposed model, we see that the LDGW distribution is the most appropriate model for analysis of the considered hydrologic data.

Table 17 Comparison via Vuong test with $Z_{Calculated}$

| Competing models	Z_{Data-I}	$Z_{Data-II}$	$Z_{Data-III}$	$Z_{Data-IV}$
LDGE-Wei.	5.2185	7.0089	8.6587	5.6671
LDGE-Fr.	2.3007	9.5239	4.6398	16.3466
LDGE-EV	4.5451	2.0263	7.7256	8.1281
LDGE-GP	− 9.3555	15.3708	3.4378	0.7066
LDGE-GW	3.6701	51.0472	12.9314	23.99929
LDGE-LN3	− 8.0081	− 1.2434	7.5549	3.8145
LDGE-Kappa3	2.2836	48.3992	− 12.3676	13.2672
LDGE-LP3	6.5921	17.0309	4.2569	11.8869
LDGE-GEEL	2.3006	4.7262	6.8553	8.5225
LDGE-GLO	3.6938	3.6939	6.6912	8.2274
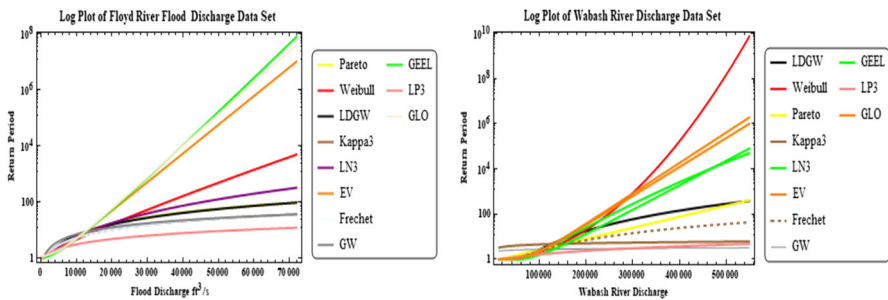
Fig. 6 Return periods of the competing models for data set-I and II

based on this observation, we further explored some characteristics of the hydrology data, which are mentioned below.

3.3.1 Return period

Flood peaks/Heavy rain fall/High temperature do not occur with any fixed pattern in time or magnitude. Time intervals between floods vary. The definition of return period is the average of these inter-event times between flood events (Rao and Hameed 2000). This implies that large floods/Heavy rain falls/High temperatures naturally have large return periods and vice versa. Such definition of the return periods may not involve any reference to probability. However, a relationship between the probability of occurrence of a flood and its return period can be justified. A given flood x_T with a return period T may be exceeded once in T years. Hence the probability of exceedance is $P(X > x_T) = \frac{1}{T}$. Now, a return level with a return period of $T = 1/p$ is a high threshold x_T (e.g., annual peak flow of a river) whose probability of exceedance is p. For this purpose the return level x_T of the LDGW distribution is given by
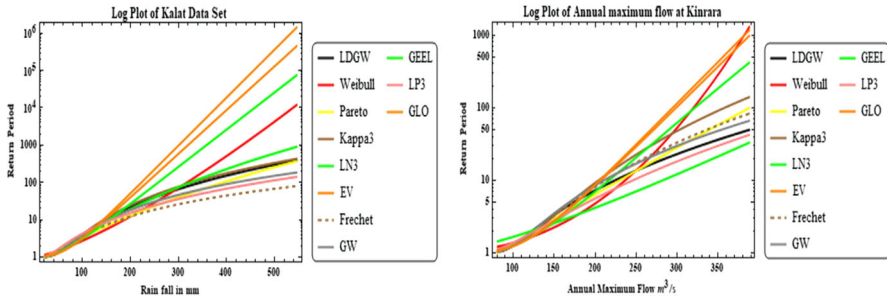
Fig. 7 Return periods of the competing models for data set-III and IV

$$
x_T = \left(\left\{ 1 - \left(\frac{\theta + \ln\left\{ (1 - e^\theta)\left(1 + \frac{1}{1 - e^\theta} - p^{-\frac{1}{\beta}}\right) \right\}}{\theta} \right)^{-\lambda} \right\} / \lambda \right)^{\frac{1}{\theta}},
$$

where $x_T > 0$ and $T \geq 1$. Table 23 provides estimates of the return level x_T for the data sets I, II, III and IV, respectively, for the return periods $T = 2, 5, 10, 25, 50, 100, 200$ years. Moreover, the return periods for some largest values of all the data sets are reported in Tables 24 and 25 and computed using $T = \frac{1}{P(x_T)}$, where $P(x_T) = S(x_T)$ is the survival function of the LDGW distribution given by

$$
S_{LDGW}(x|\lambda, \theta, \beta) = \left(1 + \frac{e^{-\theta + \theta\left(1 - \lambda x^\theta\right)^{-\frac{1}{\lambda}}} - 1}{e^\theta - 1} \right)^{-\beta},
$$

where $\hat{\theta}$, $\hat{\lambda}$ and $\hat{\beta}$ denote the MLEs of the LDGW distribution for the corresponding data set. Tables 23, 24 and 25 are given in the "Appendix C". Moreover, comparison on the basis of return period is also depicted in Figs. 6 and 7 for the above mentioned data sets. On the basis of this comparison, we can conclude that the LDGW distribution shows a significant larger and realistic (neither too large nor too short) return period as compared to competing models.

4 Conclusion

We have developed and studied a new hydrologic probability model known as the LDGW distribution, which can effectively be used in modeling the higher kurtosis data sets, and also generates realistic return return periods for all types of environmental data irrespective of any region. It can also be used in modeling the data with bathtub or upside down bathtub or increasing hazard rate shapes. Furthermore, some properties of

the proposed model are studied. Finally, the hydrological applications of the proposed model to real-life data sets demonstrate its competence, usefulness and applicability for future use. Finally, we recommend this model for data having much larger kurtosis typically greater than 20 and skewness to kurtosis ratio less than 0.29.

Appendix-A

Proof **Necessity**:

If $X \sim LDGW(\lambda, \theta, \beta)$, with a cdf defined by Equation (1), then its hrf can be expressed as

$$
h(x) = h_{LDGW}(x|\lambda, \theta, \beta) = \frac{e^{-\theta+\theta\left(1-\lambda x^\theta\right)^{-\frac{1}{\lambda}}} x^{-1+\theta} \beta\theta^2 \left(1 - x^\theta \lambda\right)^{-1-\frac{1}{\lambda}}}{\left(e^\theta - 1\right)\left(1 + \frac{e^{-\theta+\theta\left(1-\lambda x^\theta\right)^{-\frac{1}{\lambda}}}-1}{e^\theta-1}\right)},
$$

where $x > 0$, $\lambda \le 0$, $\alpha > 0$ and $\beta > 0$.

Now, differentiating the logarithmic form of the hrf with respect to x, we get

$$
\frac{d \ln h(x)}{dx} = \frac{\theta - 1}{x} + \theta^2 x^{\theta-1}\left(1 - \lambda x^\theta\right)^{-\frac{1}{\lambda}-1} + \left(\frac{1}{\lambda} + 1\right)\frac{\lambda\theta x^{\theta-1}}{1 - \lambda x^\theta}
$$

$$
- \frac{e^{-\theta+\theta\left(1-\lambda x^\theta\right)^{-\frac{1}{\lambda}}} x^{-1+\theta} \beta\theta^2 \left(1 - \lambda x^\theta\right)^{-1-\frac{1}{\lambda}}}{\left(e^\theta - 1\right)\left(1 + \frac{e^{-\theta+\theta\left(1-\lambda x^\theta\right)^{-\frac{1}{\lambda}}}-1}{e^\theta-1}\right)}, \quad \frac{h'(x)}{h(x)} - \left(\frac{1}{\lambda} + 1\right)\frac{\lambda\theta x^{\theta-1}}{1 - \lambda x^\theta}
$$

$$
= \frac{\theta - 1}{x} + \theta^2 x^{\theta-1}(1 - \lambda x^\theta)^{-\frac{1}{\lambda}-1} - \frac{e^{-\theta+\theta\left(1-\lambda x^\theta\right)^{-\frac{1}{\lambda}}} x^{-1+\theta} \beta\theta^2 \left(1 - \lambda x^\theta\right)^{-1-\frac{1}{\lambda}}}{\left(e^\theta - 1\right)\left(1 + \frac{e^{-\theta+\theta\left(1-\lambda x^\theta\right)^{-\frac{1}{\lambda}}}-1}{e^\theta-1}\right)},
$$

which after some algebraic manipulations we get Equation(8).

Sufficiency:

Suppose Equation(8) holds, then it may be re-written as

$$
\frac{h'(x)}{(h(x))^2} = \frac{\left(1 + \frac{e^{-\theta+\theta\left(1-\lambda x^\theta\right)^{-\frac{1}{\lambda}}}-1}{e^\theta - 1}\right)^{-\beta}}{\left[\beta\theta^2\left(1 + \frac{e^{-\theta+\theta\left(1-\lambda x^\theta\right)^{-\frac{1}{\lambda}}}-1}{e^\theta - 1}\right)^{-\beta-1}\frac{e^{-\theta+\theta\left(1-\lambda x^\theta\right)^{-\frac{1}{\lambda}}} x^{\theta-1}\left(1 - \lambda x^\theta\right)^{-\frac{1}{\lambda}-1}}{e^\theta - 1}\right]^2}
$$

$$
\frac{d\left\{\beta\theta^2\left(1+\frac{e^{-\theta+\theta\left(1-\lambda x^\theta\right)^{-\frac{1}{\lambda}}}-1}{e^\theta-1}\right)^{-\beta-1}\frac{e^{\theta-\theta\left(1-\lambda x^\theta\right)^{-\frac{1}{\lambda}}}x^{\theta-1}\left(1-\lambda x^\theta\right)^{-\frac{1}{\lambda}-1}}{e^\theta-1}\right\}}{dx} + 1.
$$

From the above differential equation, we have

$$
h(u) = \frac{e^{-\theta+\theta\left(1-\lambda u^\theta\right)^{-\frac{1}{\lambda}}}u^{-1+\theta}\beta\theta^2\left(1-\lambda u^\theta\right)^{-1-\frac{1}{\lambda}}}{\left(e^\theta-1\right)\left(1+\frac{e^{-\theta+\theta\left(1-\lambda u^\theta\right)^{-\frac{1}{\lambda}}}-1}{e^\theta-1}\right)}. \tag{10}
$$

Integrating the Equation (10) from 0 to x we get

$$
-\ln(1-F_{LDGW}(x|\lambda,\theta,\beta)) = -\ln\left(\left(1+\frac{e^{-\theta+\theta\left(1-\lambda x^\theta\right)^{-\frac{1}{\lambda}}}-1}{e^\theta-1}\right)^{-\beta}\right),
$$

which after simplification yields

$$
F_{LDGW}(x|\lambda,\theta,\beta) = 1 - \left(1+\frac{e^{-\theta+\theta\left(1-\lambda x^\theta\right)^{-\frac{1}{\lambda}}}-1}{e^\theta-1}\right)^{-\beta}.
$$

This completes the proof.   □

Appendix-B

$$
\frac{\partial\ell(\Theta)}{\partial\theta} = -n - \frac{e^\theta n}{e^\theta-1} + \frac{2n}{\theta} + \sum_{i=1}^n\ln(x_i) + \left(1+\frac{1}{\lambda}\right)\sum_{i=1}^n\frac{\lambda\ln(x_i)x_i^\theta}{1-\lambda x_i^\theta}
$$

$$
+ \theta\sum_{i=1}^n\ln(x_i)x_i^\theta\left(1-\lambda x_i^\theta\right)^{-1-\frac{1}{\lambda}} + \sum_{i=1}^n\left(1-\lambda x_i^\theta\right)^{-\frac{1}{\lambda}} - (1+\beta)
$$

$$
\times\sum_{i=1}^n\frac{-\frac{e^\theta\left(-1+e^{-\theta+\theta\left(1-\lambda x_i^\theta\right)^{-\frac{1}{\lambda}}}\right)}{\left(-1+e^\theta\right)^2}+\frac{e^{-\theta+\theta\left(1-\lambda x_i^\theta\right)^{-\frac{1}{\lambda}}}\left(-1+\theta\ln(x_i)x_i^\theta\left(1-\lambda x_i^\theta\right)^{-1-\frac{1}{\lambda}}-\left(1-\lambda x_i^\theta\right)^{-\frac{1}{\lambda}}\right)}{-1+e^\theta}}{1+\frac{e^{-\theta+\theta\left(1-\lambda x_i^\theta\right)^{-\frac{1}{\lambda}}}-1}{e^\theta-1}} = 0,
$$

$$
\frac{\partial\ell(\Theta)}{\partial\lambda} = \frac{1}{\lambda^2}\sum_{i=1}^n\ln(1-\lambda x_i^\theta) + \left(1+\frac{1}{\lambda}\right)\sum_{i=1}^n\frac{x_i^\theta}{1-\lambda x_i^\theta}
$$

$$
+ \theta\sum_{i=1}^n\left(1-\lambda x_i^\theta\right)^{-\frac{1}{\lambda}}\left(\frac{\ln(1-\lambda x_i^\theta)}{\lambda^2}+\frac{x_i^\theta}{\lambda\left(1-\lambda x_i^\theta\right)}\right)
$$

$$
- (1+\beta)\sum_{i=1}^n\frac{e^{-\theta+\theta\left(1-\lambda x_i^\theta\right)^{-\frac{1}{\lambda}}}\theta\left(1-\lambda x_i^\theta\right)^{-\frac{1}{\lambda}}\left(\frac{\ln(1-\lambda x_i^\theta)}{\lambda^2}+\frac{x_i^\theta}{\lambda\left(1-\lambda x_i^\theta\right)}\right)}{\left(e^\theta-1\right)\left(1+\frac{e^{-\theta+\theta\left(1-\lambda x_i^\theta\right)^{-\frac{1}{\lambda}}}-1}{e^\theta-1}\right)} = 0
$$

and

$$\frac{\partial \ell(\Theta)}{\partial \beta} = \frac{n}{\beta} - \sum_{i=1}^{n} \ln \left[1 + \frac{e^{-\theta + \theta \left(1 - \lambda x_i^{\theta}\right)^{-\frac{1}{\lambda}}} - 1}{e^{\theta} - 1} \right] = 0.$$

Appendix-C

See Tables 18,19, 20, 21, 22, 23, 24 and 25.

Table 18 Variances and co-variances of MLEs for data set-I to IV

Data set	Var($\hat{\lambda}$)	Var($\hat{\beta}$)	Var($\hat{\theta}$)	Cov($\hat{\lambda},\hat{\beta}$)	Cov($\hat{\lambda},\hat{\theta}$)	Cov($\hat{\beta},\hat{\theta}$)
I	8.1894	0.0064	0.0585	0.2597	0.6988	0.0208
II	2.3715	0.0319	0.0943	0.0606	0.1455	0.0036
III	2.9084	0.0094	0.0729	0.1921	0.4647	0.0291
IV	8.7512	0.0011	0.1935	0.1595	1.3040	0.0223

Table 19 Confidence interval with level of significance α for data set-I

Parameters	$\alpha = 0.85$	$\alpha = 0.90$	$\alpha = 0.95$	$\alpha = 0.99$
λ	$(-148.7348,$ $-140.4930)$	$(-149.3357,$ $-139.8921)$	$(-150.2229,$ $-139.0049)$	$(-151.9971,$ $-137.2307)$
θ	(12.1719,12.8684)	(12.1211,12.9193)	(12.0461,12.99426)	(11.8962,13.1442)
β	(0.3854,0.6158)	(0.3686,0.6326)	(0.3438,0.6574)	(0.2942, 0.7070)

Table 20 Confidence interval with level of significance α for data set-II

Parameters	$\alpha = 0.85$	$\alpha = 0.90$	$\alpha = 0.95$	$\alpha = 0.99$
λ	$(-693.0416,$ $-688.6065)$	$(-693.365,$ $-688.2831)$	$(-693.8424,$ $-687.8057)$	$(-694.7972,$ $-686.851)$
θ	(42.0639,42.3357)	(42.0441,42.3555)	(42.0149,42.3848)	(41.9564, 42.4433)
β	(0.4893, 0.5813)	(0.4826,0.5880)	(0.4727, 0.5979)	(0.4529, 0.61776)

Table 21 Confidence interval with level of significance α for data set-III

Parameters	$\alpha = 0.85$	$\alpha = 0.90$	$\alpha = 0.95$	$\alpha = 0.99$
λ	$(-82.2392,$ $-77.3276)$	$(-82.5973,$ $-76.9695)$	$(-83.1260,$ $-76.4408)$	$(-84.1833,$ $-75.3835)$
θ	(12.3127,13.0905)	(12.2560,13.1472)	(12.1723,13.2309)	(12.0048, 13.3984)
β	(0.3931, 0.6735)	(0.3726, 0.6939)	(0.3424, 0.7241)	(0.2821, 0.7845)

Table 22 Confidence interval with level of significance α for data set-IV

Parameters	$\alpha = 0.85$	$\alpha = 0.90$	$\alpha = 0.95$	$\alpha = 0.99$
λ	(− 296.6829, − 288.1631)	(− 297.3041, − 287.5419)	(− 298.2212, −286.6248)	(− 300.0553, − 284.7907)
θ	(42.8077,44.0748)	(42.7153,44.1672)	(42.5789,44.3036)	(42.3061,44.5764)
β	(0.1410, 0.2393)	(0.1338,0.2465)	(0.1232, 0.2571)	(0.1020, 0.2783)

Table 23 Return level estimates $\hat{x_T}$ for T of data set-I to IV

T	Return level estimate (R_I)	Return level estimate (R_{II})	Return level estimate (R_{III})	Return level estimate (R_{IV})
2	3306.69	87555.6	69.9091	122.706
5	7999.14	127456.0	110.288	174.297
10	14078.7	163280.0	147.263	224.364
25	28220.2	223722.0	210.099	308.94
50	46420.8	282447.0	271.069	389.713
100	74782.9	355340.0	346.195	487.81

Table 24 Return periods for some largest values of data set-I and II

Peak Flows (m^3/s)	Return periods	Peak flows (m^3/s)	Return periods
330	87.1178	12,600	6.4589×10^{15}
970	2542.2	60,800	9.8394×10^{23}
2940	158020	84,400	1.1277×10^{26}
4740	1.1865×10^6	160,000	3.0490×10^{30}
7170	7.7400×10^6	235,000	2.7402×10^{33}
7150	8.6393×10^6	550,000	7.0715×10^{40}

Table 25 Return periods for some largest values of data set-III and IV

Run off (mm)	Return periods	Run off (mm)	Return periods
21.7	1.0159	80.7	1.0283
58.3	1.5440	115.7	1.7321
95.8	3.6559	168.5	4.5657
105.6	4.5302	247	13.1097
133.4	7.8408	354.3	37.5181
546	391.8050	397	52.8971

References

Ahmad MI, Sinclair CD, Werritty A (1988) Log-logistic flood frequency analysis. J Hydrol 98:205–224

Akinsete A, Famoye F, Lee C (2008) The beta-Pareto distribution. Statistics 42:547–563

Alves MIG, Haan LD, Neves C (2009) Statistical inference for heavy and super-heavy tailed distributions. J Stat Plan Inf 139:213–227

Balakrishnan N, Leung MY (1988) Means, variances and covariances of order statistics, BLUEs for the Type-I generalized logistic distribution, and some applications. Communications in Statistics Simulation and Computation 17(1):51–84

Bobee B (1975) The log Pearson type 3 distribution and its application in hydrology. Water Resour Res 11(5):681–689

Cakmakyapan S, Ozel G (2016) The Lindley family of distributions: properties and applications. Hacet J Math Stat 46:1–27

Dargahi-Noubary GR (1989) On tail estimation: an improved method. Math Geol 21(8):829–842

Denuit M, Maréchal X, Pitrebois S, Walhin JF (2007) Actuarial modelling of claim counts risk classification, credibility and bonus-malus systems. Wiley, West Sussex

Dyrrdal AV (2012) Estimation of extreme precipitation in Norway and a summary of the state of the art. Report no. 08/2012, Climate, Norwegian Meteorological Institute

Foss S, Zachary S, Korshunov D (2011) An introduction to heavy-tailed and sub-exponential distributions. Springer, New York

Hosking JRM, Wallis JR (1987) Parameter and quantile estimation for the generalized Pareto distribution. Technometrics 29(3):339–349

Hussain T, Bakouch HS, Iqbal Z (2018) A new probability model for hydrologic events: properties and applications. J Agric Biol Environ Stat 23(1):63–82

IPCC (2012) Managing the risks of extreme events and disasters to advance climate change adaptation. Field CB et al (eds) Cambridge University Press

Johnson NL, Kotz S, Balakrishnan N (2005) Continuous univariate distributions, vol 1, 2nd edn. Wiley, New York

Junior PWM, Johnson ES (1973) Three parameter kappa distribution maximum likelihood estimates and likelihood ratio tests. Mon Weather Rev 101(09):701–711

Krige D (1960) On the departure of ore value distributions from the log-normal model in South African gold mines. J South Afr Inst Min Metall 40(1):231–244

Leadbetter MR, Lindgren G, Rootzen H (1987) Extremes and related properties of random sequences and processes. Springer, NewYork

Loikith PC, Neelin JD (2015) Short-tailed temperature distributions over North America and implications for future changes in extremes. Geophys Res Lett 42:8577–8585

Lomax KS (1987) Business failures: another example of the analysis of failure data. J Am Stat Assoc 49:847–852

Markovich N (2007) Nonparametric analysis of univariate heavy-tailed data. Wiley, Chichester

Mudholkar GS, Srivastava DK, Kollia GD (1996) A generalization of the Weibull distribution with application to the analysis of survival data. J Am Stat Assoc 91(436):1575–1583

Mujere N (2011) Flood frequency analysis using the Gumbel distribution. Int J Comput Sci Eng 3(7):2774–2778

Papalexiou SM, Koutsoyiannis D, Makropoulos C (2013) How extreme is extreme? An assessment of daily rainfall distribution tails. Hydrol Earth Syst Sci 17:851–862

Pickands J (1975) Statistical inference using extreme order statistics. Ann Stat 3:119–131

R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/

Ramos PL, Louzada F, Ramos E, Dey S (2018) The Fréchet distribution: estimation and application an overview. Available at: arXiv:1801.05327v1 [stat.AP]

Rao AR, Hameed KH (2000) Flood frequency analysis: new directions in civil engineering. CRC Press, Florid

Vuong QH (1989) Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57(2):307–333

Tassaddaq Hussain received his Ph.D. degree in Applied Statistics from National College of Business Administration and Economics (NCBA&E) Lahore, Pakistan under HEC Pakistan indigenous scholarship scheme in 2012. Since 2018, he has been a permanent Faculty member as Assistant Professor of Mirpur University of Science and Technology (MUST) Mirpur (AJK), Pakistan. His research activities are focused on Characterization of probability models and distribution theory with applications to real-life data.

Hassan S. Bakouch received his Ph.D. degree in Mathematical Statistics from Tanta University in 2006. From 2006 to 2014, he was an Assistant Professor at Tanta University. Since 2014, he is an Associate Professor of Mathematical Statistics in the Department of Mathematics at Tanta University. His research activities are focused on time series modeling and distribution theory with applications to real-life data. Also, his research includes estimation of population abundance via detection functions for line transect data. He published more fifty research papers in different refereed and impacted journals. Further, he has active collaborations with researchers in different countries.

Christophe Chesneau received the Ph.D. degree in Applied Mathematics from the University of Paris VI in 2006. Since 2007, he is an Assistant Professor in the Department of Mathematics at the University of Caen-Normandie. His research activities are focused on the applied mathematics, statistics and probability with applications.