

A New Method for Generating Distributions: An Application to Flow Data

Hüseyin Ünözkan^{1,*}, Mehmet Yilmaz²

¹Ankara University, Graduate School of Natural and Applied Science, Department of Statistics, Turkey

²Ankara University, Science Faculty, Department of Statistics, Turkey

Abstract Nowadays, one of the aim of statistical studies is to provide the future with the models which are easily accessible and simple. Therefore, more suitable distributions are needed to model the data. In this study, a new distribution is generated with exponential marginals Farlie-Gumbel-Morgenstern distribution. Specifications and characteristics of this new distribution are studied. The structure of the proposed distribution is discussed statistically and the parameter estimation for the new distribution is made by known methods. In addition, reliability analysis has performed. Due to the shape and flexibility of the proposed distribution, it is thought to be an alternative to distributions which are used for modeling flow data. Efficiency on the statistical modeling of the new distribution can be detected by using flow data sets in literature. Furthermore, Terme and Sefaatli Creeks' flow-data obtained from Turkish State Water Affairs Directorate are used to model. It is concluded that this new distribution offers a model that can be used effectively in stream flows.

Keywords Farlie-Gumbel-Morgenstern Distribution, Copula, Reliability Analysis, Generating Distribution, Exponential Distribution

1. Introduction

Global warming has effected nearly every water sources. For estimating behavior of streams, modeling and analyzing flows has vital importance. In the world, the population rate is rising rapidly, which makes pure water resources more important. For some decades, researchers have studied in streams for estimating drought and flood. Till now, they have used some most known distributions for modeling streams statistically. After modeling, they have analyzed behavior of streams easily.

Some stream statistics was described as magnitude, variability and flow extremes. Mean flow is the fundamental statistics of flow record. It is usually expressed as flow in m³/s. In order to measure the variability, it is the most common method to obtain the distribution of monthly average flows [3].

Therefore, we can use the mean of annual flows, mean of the lowest month flows, and rivers' total flows to model their behavior. For low flows, Generalized Extreme Value distribution, Pearson Type III and Generalized Logistic distributions are commonly used. Sometimes these

distributions can be used with more parameters than their original structure. A fourth distribution, the Generalized Pareto, was also offered [11]. In literature, we can find some transmutations for these distributions to better match streams flow data [4].

[11] shows which distributions are needed to analyze flow data, and in particular to analyze minimum values. We can estimate the behavior of the stream with mean flows, such that this can be important for living beings in the district of this stream.

The frequency of floods was analyzed in [7] by hydrological functions through extremes and they used the same statistical distributions in [11]. In this paper, we suggest a new distribution for analyzing every kinds of stream data which are using for describing river behaviors.

Through the paper, this proposed distribution is called as "CFGMWEM". CFGMWEM presents good results especially in modeling low flows. In our presentation, we have first introduced CFGMWEM. After this we have shown new distribution's properties, and important characteristics. Finally, we have compared CFGMWEM with most known hydrologic statistical functions via data which were used in literature and new data obtained from Turkey rivers.

2. Materials and Methods [New Distribution (CFGMWEM)]

In recent years, during the measuring process of flows,

* Corresponding author:

hunozkan@gmail.com (Hüseyin Ünözkan)

Published online at <http://journal.sapub.org/statistics>

Copyright © 2019 The Author(s). Published by Scientific & Academic Publishing

This work is licensed under the Creative Commons Attribution International

License (CC BY). <http://creativecommons.org/licenses/by/4.0/>

mostly logical systems have been used. While measuring these values, logical machines obtain observation only if the flow level is higher than the lowest measuring point of them. According to this approach for fertile measurement, the heaviness of stream has to be more than a fixed point.

The aim of this study is to develop a statistical model about the flow data which is observed during this real measurement process. In order to realize this, it is necessary to obtain a successful statistical model and this requires the use of an important theorem.

Theorem (Sklar's Theorem)

Let F be a joint cumulative distribution function and H and G is marginals, then there is a copula function C in \mathbb{R} for every x and y [9].

$$F(x, y) = C(H(x), G(y))$$

Farlie-Gumbel-Morgenstren (FGM) copula with marginals u and v can be written as below [6].

$$C_{\theta}(u, v) = uv + \theta uv(1 - u)(1 - v)$$

Hence, Two dimensional Bivariate FGM distribution with marginals $H(x)$ and $G(y)$ is as follows.

$$F(x, y) = H(x)G(y)[1 + \lambda \bar{H}(x)\bar{G}(y)].$$

Probability density function of this distribution is as below.

$$f(x, y) = h(x)g(y)[1 + \lambda(1 - 2H(x))(1 - 2G(y))]$$

Under $Y = y$ condition, X has a conditional probability density function as follows.

$$\frac{f(x, y)}{g(y)} = h(x)[1 + \lambda(1 - 2H(x))(1 - 2G(y))]$$

Under $Y = y$ condition, X has a conditional distribution as below.

$$Pr(X \leq x|Y = y) = \int h(x)[1 + \lambda(1 - 2H(x))(1 - 2G(y))] dx$$

$$Pr(X \leq x|Y = y) = H(x) - \lambda(1 - 2G(y))H(x)\bar{H}(x)$$

Under $T = t$ condition probability of $X \leq t$ is

$$Pr(X \leq t|Y = t) = H(t) - \lambda(1 - 2G(t))H(t)\bar{H}(t)$$

provided.

Considering the models related to natural events, exponential distribution has a wide range of usability. Because of simple statistical structure and memoryless property exponential distribution has been used widely.

Suppose that $(t) = G(t) = 1 - e^{-\theta t}$. Then we have

$$F(t) = (1 + \lambda)H(t) - \lambda H(t)^2(3 - 2H(t))$$

We know from the literature that the transmuted distribution with baseline $H(t)$ is $(1 + \lambda)H(t) - \lambda H(t)^2$. Here, $H(t)^2$ is the failure distribution of the two-component parallel system (with identical and independent) namely, represented as $H_{2:2}$. In the light of this idea, $F(t)$ can be also rewritten as the form of $(1 + \lambda)H(t) - \lambda H_{3:2}(t)$ where $H_{3:2}$ represents a failure distribution of 3 out of 2 system with independent and identical component. Thus, we have a different form of

transmuted distribution. Hence when baseline distribution is assumed to be exponential we have the following special form of distribution.

$$F(t) =$$

$$(1 + \lambda)(1 - e^{-\theta t}) - \lambda(1 - e^{-\theta t})^2(3 - 2(1 - e^{-\theta t})) \quad (1)$$

Probability density function of CFGMWEM is as below.

$$f(t) = \frac{d}{dt}(1 + \lambda)(1 - e^{-\theta t}) - \lambda(1 - e^{-\theta t})^2(3 - 2(1 - e^{-\theta t}))$$

$$f(t) = (\theta e^{-\theta t})(1 + \lambda - 6\lambda e^{-\theta t} + 6\lambda e^{-2\theta t}),$$

$$\lambda \in [-1, 1], \theta > 0 \quad (2)$$

Some shapes of probability density function are as below. In Fig. 1 and Fig. 2 $\theta = 0.01$,

Table 1. Values of λ parameter matching by colour in Fig. 1

Colour	Blue	yellow	green	black	Red
Value	0.9	0.75	0.25	0	-0.25

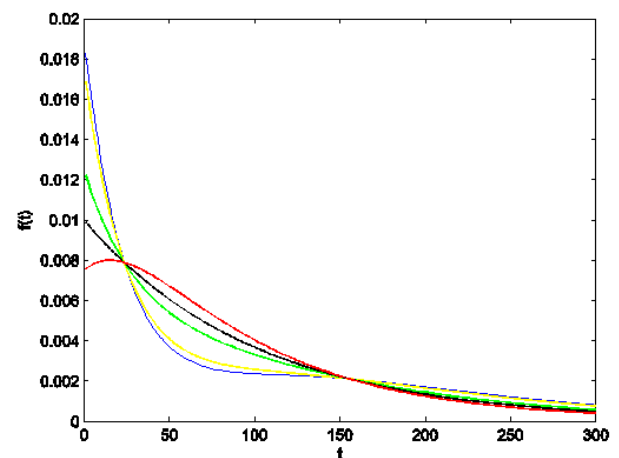


Figure 1. Plots of probability density function

Table 2. Values of λ parameter matching by colour in Fig. 2

Colour	Blue	yellow	green	black	Red
Value	-0.9	-0.75	-0.25	0	0.25

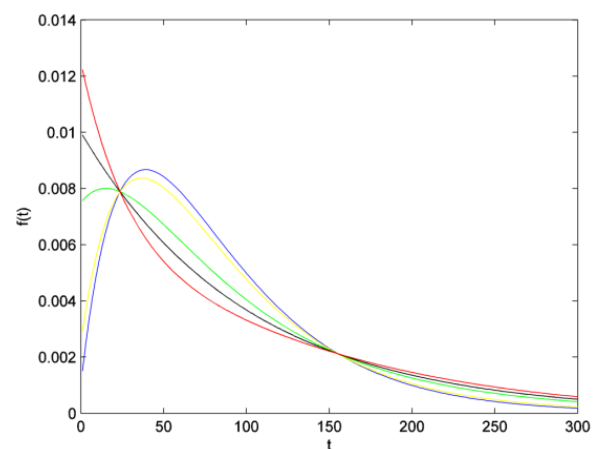


Figure 2. Plots of probability density function

Survival function of CFGMWEM is as follows;

$$\begin{aligned} S(t) &= 1 - F(t) \\ &= 1 - (1 + \lambda)(1 - e^{-\theta t}) \\ &\quad + \lambda(1 - e^{-\theta t})^2(3 - 2(1 - e^{-\theta t})) \\ &= 1 - (1 + \lambda)(H(t)) + \lambda(H(t))^2(3 - 2(H(t))) \\ S(t) &= (1 + \lambda)\bar{H}(t) - \lambda(3(\bar{H}(t))^2 - 2(\bar{H}(t))^3) \end{aligned}$$

Hazard rate function of CFGMWEM is as below.

$$\begin{aligned} r(t) &= \frac{f(t)}{S(t)} \\ &= \frac{(\theta e^{-\theta t})(1 + \lambda - 6\lambda e^{-\theta t} + 6\lambda e^{-2\theta t})}{1 - (1 + \lambda)(1 - e^{-\theta t}) + \lambda(1 - e^{-\theta t})^2(3 - 2(1 - e^{-\theta t}))} \\ &= \frac{\theta(1 + \lambda - 6\lambda e^{-\theta t} + 6\lambda e^{-2\theta t})}{(1 + \lambda - 3\lambda e^{-\theta t} + 2\lambda e^{-2\theta t})} \end{aligned}$$

If we want to calculate the risk of the starting point;

$$\lim_{t \rightarrow 0} \left(\frac{\theta(1 + \lambda - 6\lambda e^{-\theta t} + 6\lambda e^{-2\theta t})}{(1 + \lambda - 3\lambda e^{-\theta t} + 2\lambda e^{-2\theta t})} \right) = \theta(1 + \lambda)$$

We reach that in the long term risk, there are two different results according to the parameter λ .

$$\lim_{t \rightarrow \infty} \left(\frac{\theta(1 + \lambda - 6\lambda e^{-\theta t} + 6\lambda e^{-2\theta t})}{(1 + \lambda - 3\lambda e^{-\theta t} + 2\lambda e^{-2\theta t})} \right) = \theta, \lambda \neq -1$$

$$\lim_{t \rightarrow \infty} \left(\frac{6\theta(1 - e^{-\theta t})}{(3 - 2e^{-\theta t})} \right) = 2\theta, \lambda = -1$$

Some shapes of hazard rate function are as below. In Fig. 3 and Fig. 4 $\theta = 0.1$

Table 3. Values of λ parameter matching by colour in Fig. 3

Colour	Blue	yellow	green	black	Red
Value	0.9	0.75	0.25	0	-0.25

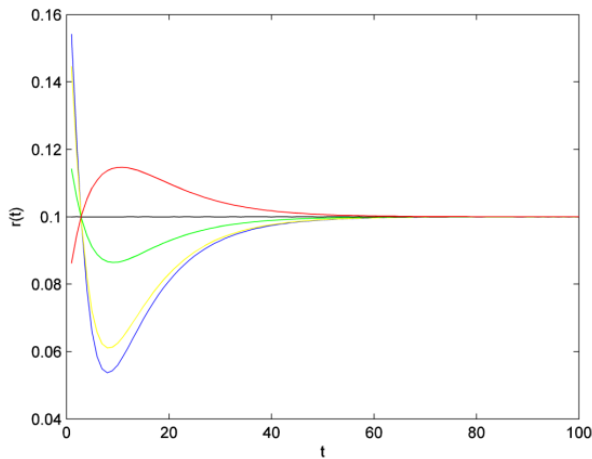


Figure 3. Plots of hazard rate function

Table 4. Values of λ parameter matching by colour in Fig. 4

Colour	red	blue	yellow	green	black	Red
Value	-1	-0.9	-0.75	-0.25	0	0.25

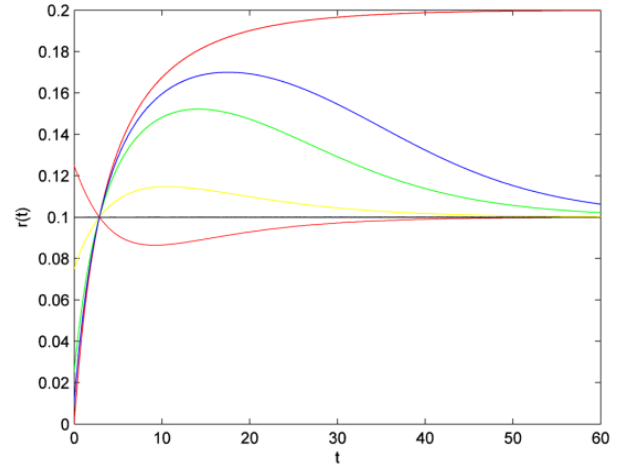


Figure 4. Plots of hazard rate function

According to Figure 3 and Figure 4, we can easily see that parameter λ changes both the shape of probability density function and hazard rate function. Here, we believe that CFGMWEM can be used in interesting data groups that may pose two different types of risks.

When parameter λ has a value between $[-1, 0)$, hazard rate function becomes bathtub shape. However, there are initial decreasing proportions of deaths, and at the beginning some components rapidly deteriorate. Thereafter a balance is formed and an almost constant hazard rate is observed. In the last part of the curve, components that complete their lifetimes are increasingly out of the system and the life span of the system is completed.

When parameter λ has a value between $(0, 1]$, hazard rate function becomes the opposite of bathtub shape. This curve is symmetric to value of parameter θ which is the hazard rate of exponential distribution. At the beginning, there are high risk, and initially some components are rapidly deteriorating. Thereafter a balance is formed and an almost constant hazard rate is observed. In the last chapter, components that complete their life span are decreasingly out of the system and the life span of the system is completed. This shape is called upside-down bathtub, inverse bathtub or unimodel.

2.1. Characteristics of Distribution

Moment generating function

$$\begin{aligned} E(e^{vX}) &= M_X(v) \\ &= \int e^{vx} f(x) dx \\ &= \int_0^\infty e^{vx} (\theta e^{-\theta x})(1 + \lambda - 6\lambda e^{-\theta x} + 6\lambda e^{-2\theta x}) dx \\ &= \int_0^\infty (\theta e^{-x(\theta-v)}(1 + \lambda) - 6\lambda \theta e^{-x(2\theta-v)} \\ &\quad + 6\lambda \theta e^{-x(3\theta-v)}) dx \\ &= (1 + \lambda) \int_0^\infty \theta e^{-x(\theta-v)} dx - 6\lambda \int_0^\infty \theta e^{-x(2\theta-v)} dx \\ &\quad + 6\lambda \int_0^\infty \theta e^{-x(3\theta-v)} dx \end{aligned}$$

$$M_X(v) = \frac{\theta(1+\lambda)}{\theta-v} - \frac{6\lambda\theta}{2\theta-v} + \frac{6\lambda\theta}{3\theta-v}, \theta > v$$

This is a linear combination of exponential distribution with $\frac{1}{\theta}$, $\frac{1}{2\theta}$ and $\frac{1}{3\theta}$ mean. Through this, for $Y_j \sim \text{Exponential}\left(\frac{1}{j\theta}\right)$, $j = 1, 2, 3$ moment generating function can be written as $M_{Y_j}(v) = \frac{j\theta}{j\theta-v}$

$$M_X(v) = (1+\lambda)M_{Y_1}(v) - 3\lambda M_{Y_2}(v) + 2\lambda M_{Y_3}(v), \theta > v \quad (3)$$

k. th pure moment

Pure moments can be easily achieved with moment generating function.

$$\begin{aligned} E(X^k) &= \left. \frac{d^k}{dv^k} M_X(v) \right|_{v=0} \\ &= (1+\lambda) \left. \frac{d^k}{dv^k} M_{Y_1}(v) \right|_{v=0} - 3\lambda \left. \frac{d^k}{dv^k} M_{Y_2}(v) \right|_{v=0} \\ &\quad + 2\lambda \left. \frac{d^k}{dv^k} M_{Y_3}(v) \right|_{v=0} \\ &= (1+\lambda) \frac{\Gamma(k+1)}{\theta^k} - 3\lambda \frac{\Gamma(k+1)}{2^k \theta^k} + 2\lambda \frac{\Gamma(k+1)}{3^k \theta^k} \\ E(X^k) &= \frac{k!}{\theta^k} \left[(1+\lambda) - 3\lambda \frac{1}{2^k} + 2\lambda \frac{1}{3^k} \right] \end{aligned}$$

Expected value

$$\begin{aligned} E(X) &= \int x f(x) dx \\ &= \int_0^\infty x (\theta e^{-\theta x}) (1+\lambda - 6\lambda e^{-\theta x} + 6\lambda e^{-2\theta x}) dx \\ &= (1+\lambda) \int_0^\infty x \theta e^{-\theta x} dx - 3\lambda \int_0^\infty 2x \theta e^{-2\theta x} dx \\ &\quad + 2\lambda \int_0^\infty 3x \theta e^{-3\theta x} dx \\ E(X) &= \frac{\lambda+1}{\theta} - \frac{3\lambda}{2\theta} + \frac{2\lambda}{3\theta} = \frac{6+\lambda}{6\theta} \\ E(X^2) &= \int x^2 f(x) dx \\ &= \int_0^\infty x^2 (\theta e^{-\theta x}) (1+\lambda - 6\lambda e^{-\theta x} + 6\lambda e^{-2\theta x}) dx \\ &= (1+\lambda) \int_0^\infty x^2 \theta e^{-\theta x} dx - 3\lambda \int_0^\infty 2x^2 \theta e^{-2\theta x} dx \\ &\quad + 2\lambda \int_0^\infty 3x^2 \theta e^{-3\theta x} dx \\ E(X^2) &= \frac{2(\lambda+1)}{\theta^2} - \frac{6\lambda}{4\theta^2} + \frac{4\lambda}{9\theta^2} = \frac{36+17\lambda}{18\theta^2} \end{aligned}$$

Variance

$$\begin{aligned} \text{Var}(X) &= E(X^2) - E(X)^2 = \frac{36+17\lambda}{18\theta^2} - \left(\frac{6+\lambda}{6\theta} \right)^2 \\ &= \frac{36+22\lambda-\lambda^2}{36\theta^2} \end{aligned}$$

2.2. Moment Estimator

We find moment estimator with matching sample moments and pure moments of distribution. By this matching we can reach moment estimation under a condition

$$\frac{38}{25} \leq \frac{m_2}{m_1^2} \leq \frac{106}{49}.$$

$$\hat{\theta} = \frac{102m_1 - \sqrt{10404m_1^2 - 4752m_2}}{36m_2}$$

$$\begin{aligned} \hat{\lambda} &= m_1 6 \left(\frac{102m_1 - \sqrt{10404m_1^2 - 4752m_2}}{36m_2} \right) - 6, \frac{38}{25} \\ &\leq \frac{m_2}{m_1^2} \leq \frac{106}{49} \end{aligned}$$

2.3. Maximum Likelihood Estimation

$$\begin{aligned} L(\theta, \lambda; \mathbf{t}) &= f(t_1, t_2, t_3, \dots, t_n; \theta, \lambda) = \prod_{i=1}^n f(t_i; \theta, \lambda) \\ &= \prod_{i=1}^n \theta e^{-\theta t_i} (1+\lambda - 6\lambda e^{-\theta t_i} + 6\lambda e^{-2\theta t_i}) \\ &= \theta^n e^{-\theta \sum_{i=1}^n t_i} \prod_{i=1}^n (1+\lambda - 6\lambda e^{-\theta t_i} + 6\lambda e^{-2\theta t_i}) \end{aligned}$$

With using Log Likelihood, maximum likelihood estimation of parameters can be obtained with derivation of λ and θ .

$$\begin{aligned} \log(L(\theta, \lambda; \mathbf{t})) &= n \log \theta - \theta \sum_{i=1}^n t_i \\ &\quad + \sum_{i=1}^n \log(1+\lambda - 6\lambda e^{-\theta t_i} + 6\lambda e^{-2\theta t_i}) \\ \frac{\partial}{\partial \lambda} \log(L(\theta, \lambda; \mathbf{t})) &= \sum_{i=1}^n \frac{1 - 6e^{-\theta t_i} + 6e^{-2\theta t_i}}{1+\lambda - 6\lambda e^{-\theta t_i} + 6\lambda e^{-2\theta t_i}} = 0 \\ \frac{\partial}{\partial \theta} \log(L(\theta, \lambda; \mathbf{t})) &= \frac{n}{\theta} - \sum_{i=1}^n t_i + \sum_{i=1}^n \frac{t_i 6\lambda e^{-\theta t_i} - t_i 12\lambda e^{-2\theta t_i}}{1+\lambda - 6\lambda e^{-\theta t_i} + 6\lambda e^{-2\theta t_i}} = 0 \end{aligned}$$

2.4. L Moments Estimation

In l moments estimation we equalize l samples to l moments. Below we will first give l samples and the calculations of these characteristics. After this we will give l moments and their calculations. At last we will match them and find the estimations for parameters.

Now we can express l samples as follows. $x_{(j)}$ shows order statistics.

$$l_1 = \frac{1}{n} \sum_{j=1}^n x_{(j)}$$

$$l_2 = \frac{1}{n} \sum_{j=2}^n \frac{j-1}{n-1} x_{(j)}$$

$$L_1 = l_1$$

$$L_2 = 2l_2 - l_1$$

$$\lambda_1 = E(X)$$

$$\lambda_2 = \frac{E(X_{2:2}) - E(X_{1:2})}{2}$$

There are two equations with l samples and l moments matching.

$$L_1 = \lambda_1$$

$$L_2 = \lambda_2$$

By using these two equations and adjusting them to CFGMWEM, the equations and matchings are observed as follows.

$$\lambda_1 = E(X) = \frac{6 + \lambda}{6\theta}$$

$$E(X_{2:2}) = \frac{20\lambda + 60\theta - \lambda^2 + 90}{60\theta}$$

$$E(X_{1:2}) = \frac{60\theta + \lambda^2 + 30}{60\theta}$$

$$\lambda_2 = \frac{E(X_{2:2}) - E(X_{1:2})}{2} = \frac{10\lambda - \lambda^2 + 30}{60\theta}$$

$$\frac{\lambda_2}{\lambda_1} = \frac{L_2}{L_1}$$

$$\frac{2l_2 - l_1}{l_1} = \frac{\frac{10\lambda - \lambda^2 + 30}{60\theta}}{\frac{6 + \lambda}{6\theta}}$$

$$\frac{\lambda^2 + 30}{10\lambda + 60} = 2 \left(1 - \frac{l_2}{l_1} \right), \text{ in prospect of } 2 \left(1 - \frac{l_2}{l_1} \right) = c$$

$$\lambda = 5c - 5\sqrt{c^2 - 1.2 + 2.4c}$$

$$\text{For } \lambda \in [-1, 1], c \text{ have to be in } \left[\frac{31}{50}, \frac{31}{70} \right]$$

Now parameter estimation methods may be evaluated via examining Table 5. In Table 5 there are estimation results of three methods. Numerical technique is used for the results and repeating number is 100 for every observation. n shows the observation number. Rmse is root mean squares of error.

Table 5. Parameter estimation results

$\lambda = -0.9 \theta = 0.05$ Moment Estimation				
	n=10	n=50	n=100	n=1000
$\hat{\lambda}$	-0.5958	-0.7357	-0.8494	-0.8949
Rmse($\hat{\lambda}$)	0.5103	0.3243	0.1412	0.0683
$\hat{\theta}$	0.0571	0.0518	0.0504	0.0500
Rmse($\hat{\theta}$)	0.0167	0.0056	0.0043	0.0014
$\lambda = -0.9 \theta = 0.05$ Maximum Likelihood Estimation				
	n=10	n=50	n=100	n=1000
$\hat{\lambda}$	-0.8731	-0.8673	-0.8802	-0.9064

Rmse($\hat{\lambda}$)	0.2905	0.1942	0.1409	0.0434
$\hat{\theta}$	0.0532	0.0508	0.0503	0.0500
Rmse($\hat{\theta}$)	0.0138	0.0059	0.0035	0.0013

$\lambda = -0.9 \theta = 0.05$ L Moments Estimation				
	n=10	n=50	n=100	n=1000
$\hat{\lambda}$	-0.4058	-0.7417	-0.8121	-0.9000
Rmse($\hat{\lambda}$)	0.7058	0.2611	0.1679	0.0544
$\hat{\theta}$	0.0592	0.0511	0.0511	0.0499
Rmse($\hat{\theta}$)	0.0170	0.0060	0.0044	0.0010

$\lambda = 0.2 \theta = 5$ Moment Estimation				
	n=10	n=50	n=100	n=1000
$\hat{\lambda}$	-0.4545	-0.1828	-0.0550	0.2014
Rmse($\hat{\lambda}$)	0.8239	0.5976	0.5049	0.2456
$\hat{\theta}$	4.9762	4.7212	4.9559	4.9872
Rmse($\hat{\theta}$)	1.4727	0.8628	0.7305	0.2411

$\lambda = 0.2 \theta = 5$ Maximum Likelihood Estimation				
	n=10	n=50	n=100	n=1000
$\hat{\lambda}$	0.1918	0.1843	0.1692	0.1892
Rmse($\hat{\lambda}$)	0.7293	0.3856	0.2413	0.0750
$\hat{\theta}$	5.4206	4.9986	5.0843	4.9991
Rmse($\hat{\theta}$)	2.0231	0.7746	0.6097	0.1532

$\lambda = 0.2 \theta = 5$ L Moments Estimation				
	n=10	n=50	n=100	n=1000
$\hat{\lambda}$	-0.0994	0.1839	0.1384	0.2084
Rmse($\hat{\lambda}$)	0.5938	0.3997	0.3292	0.1293
$\hat{\theta}$	5.3078	5.0657	5.0417	5.0007
Rmse($\hat{\theta}$)	2.1408	0.8289	0.5662	0.1927

According to Table 5, maximum likelihood estimation is better than other two estimation methods. Because of this maximum likelihood estimation will be used in application part.

3. Results and Discussion

Now, using some different flow data, we first compare CFGMWEM with the most common hydrological statistical distributions. Subsequently, we offer CFGMWEM as a new distribution for flow data with different kinds of data groups. While comparing distributions, we will use Kolmogorov-Smirnov test statistics for looking the availability of our distribution to data sets. In Kolmogorov-Smirnov test statistics p value indicates the success rate of distribution in the explanation. [5], [8]

Once we see that the two distributions are equal, we will have a new problem that which distribution is better for this data set. Because according to the hypothesis test, there may be many distributions that equal to nonparametric

distribution. Akaike Information Criterion (AIC) can be used to compare these distributions. When AIC is used, the distribution with the minimum AIC value is selected as the best distribution. Since the AIC is a penalty value and the minimum value represents the maximum similarity to the non-parametric distribution of the data set, the minimum AIC value is the maximum similarity to the distribution. [1], [10]

In this section, CFGMWEM will be compared with most known flow distributions using some different flow data. While comparing distributions, Kolmogorov-Smirnov test statistics will be used. When using Kolmogorov-Smirnov statistics, the least statistical value is considered to be the best modeling. p value of Kolmogorov-Smirnov statistics informs us about plausibility of the conformity.

Data 1: The first data we used are the flood peak values (in m^3/s) of the Wheaton River near Carcross in Yukon Territory, Canada. The data consist of 72 exceedances for the years 1958–1984, rounded to one decimal place. This data was analyzed in [2] and after this the same data was used in [4].

Table 6. Wheaton River Flood Peaks (m^3/s) data

1.7	2.2	14.4	1.1	0.4	20.6	3.4
12	9.3	1.4	18.7	8.5	25.5	11.9
1.1	2.5	14.4	1.7	37.6	0.6	21.5
15	11	7.3	22.9	1.7	0.1	5.3
1.7	7	20.1	0.4	14.1	9.9	2.5
3.6	5.6	30.8	13.3	4.2	25.5	2.8
5.3	0.7	13	27.6	64	27.4	27.1
11.6	14.1	22.1	20.2	9.7	27	2.5
2.2	39	0.3	36.4	1.5	1	2.7
1.1	0.6	9	16.8	27.5	1.9	30
10.4	10.7					

Table 7. Wheaton River Flood Peaks (m^3/s) data test results

Model	K-S	p
GEV	0.1398	0.1087
Log Pearson III	0.09948	0.44568
Generalized Pareto	0.1005	0.43177
Weibull	0.09045	0.5666
Log-Normal	0.1394	0.1103
CFGMWEM	0.0827	0.6774

In terms of Table 7 CFGMWEM presents best model and clarifies data more than other distributions.

Data 2: This data group is the values of Terme Creek's total flows in March from 1969 to 2000. The data group was received from Turkish State Water Affairs Directorate. In introduction section we told that there were two kinds of flow values that we looked for. These values are extremes and low flows. Therefore, we use the Terme Creek's total flows in March for extreme values analysis because it takes the highest values in March.

Table 8. Terme Creek's Total Flows (hm^3) in March data

72.4	39.4	15.6	19.1	19.1	3.93	28.5
10.8	43.4	77	23.1	22.5	25.5	14.1
19.34	7.24	19	4.55	23.8	20.4	3.95
24.6	17.5	20.4	23.6	11.3	13.4	27.2
10	23.9	2.09				

Table 9. Terme Creek's Total Flows (hm^3) in March test results

Model	K-S	p
GEV	0.1419	0.5148
Log Pearson III	0.1490	0.4530
Generalized Pareto	0.1515	0.4319
Weibull	0.13766	0.5533
Log-Normal	0.17121	0.2892
CFGMWEM	0.1418	0.5158

In reference to Table 9, CFGMWEM presents second best model and clarifies data more than the other four distributions.

Data 3: This data group is the values of Terme Creek's total flows in September from 1969 to 2000. In introduction section we told that there were two kinds of flow values that we looked for. These values are extremes and low flows. Therefore, we use the Terme Creek's total flows in September for low values analysis because it takes the lowest values in September.

Table 10. Terme Creek's Total Flows (hm^3) in September data

2.55	2.3	4.21	2.57	0.26	0.44	0.63
1.07	1.04	1.35	2.67	1.98	1.71	0.14
0.494	0.103	0.758	0.077	0.218	0.099	0.011
4.48	1.78	1.68	0.831	0.299	0.3	0.134
0.28	0.37	0.266				

Table 11. Terme Creek's Total Flows (hm^3) in September test results

Model	K-S	p
GEV	0.1402	0.5300
Log Pearson III	0.0967	0.9072
Generalized Pareto	0.1057	0.8429
Weibull	0.1035	0.8604
Log-Normal	0.1102	0.8060
CFGMWEM	0.0773	0.9855

Given the results of Table 11 of the CFGMWEM, it is clear that CFGMWEM offers good models and can be used to estimate the data related to flows, as well as other distributions known as flow distributions.

Data 4: This data group contains 56 measurements of total flows from Sefaatli Creek in April from 1953 to 2014. The data group was received from Turkish State Water Affairs Directorate. We use the Sefaatli Creek's mean flows in April for extremes analysis because it takes the highest values in April.

Table 12. Sefaati Creek's Mean Flows (m^3/s) in April data

29	44.9	9.5	12.5	10.8	19.8	18.3
29.6	61.7	46.8	41.8	13.6	17.8	15.1
26.9	14.1	36.3	23	18.4	19.5	41.5
27.2	25.2	22	39.6	4.94	16.2	31.4
6.21	2.99	4.12	2.47	19.99	9.844	5.525
9.23	6.89	21.6	12.2	12.9	17.3	27.6
7.72	8.14	29	36.5	31	4.248	3.52
44.4	15.2	44.5	53.4	23.4	5.19	25.8

Table 13. Sefaati Creek's Mean Flows (m^3/s) in April test results

Model	K-S	p
GEV	0.0545	0.9930
Log Pearson III	0.0548	0.9926
Generalized Pareto	0.0495	0.9980
Weibull (3P)	0.0500	0.9977
Log-Normal (3P)	0.0626	0.9705
CFGMWEM	0.0536	0.9944

Data 5: This data group is the values of Sefaati Creek's total flows in August from 1953 to 2005. The data group was received from Turkish State Water Affairs Directorate. In introduction section we told that there were two kinds of flow values that we looked for. These values are extremes and low flows. Therefore, we use the Sefaati Creek's total flows in August for low values analysis because it takes the lowest values in August.

Table 14. Sefaati Creek's Mean Flows (m^3/s) in August data

0.92	5.9	0.28	0.25	0.21	0.8	0.6
4.37	5.96	2.32	0.12	1.03	0.583	0.01
1.22	1.51	0.353	0.476	1.97	1.59	0.31
1.06	3.66	1.55	2.54	1.79	3.49	4.92
0.15	0.013	0.943	0.139	0.712	4.6	3.03
0.858	0.63	1.29	0.696	0.052	3.21	3.38
0.626	4.23	0.371	2.8	0.864	5.99	2.53
4.92						

Table 15. Sefaati Creek's Mean Flows (m^3/s) in August test results

Model	K-S	p
GEV	0.1211	0.3547
Log Pearson III	Not Fit	
Generalized Pareto	0.1026	0.5612
Weibull (3P)	0.1403	0.2004
Log-Normal (3P)	0.0915	0.7008
CFGMWEM	0.0617	0.9852

In terms of Table 13 and Table 15, results point out the same conclusion with results of Terme Creek. Especially at low flows, CFGMWEM offers better modeling than other most common flow distributions.

4. Conclusions

But how CFGMWEM conform with both extremes and low flows in the same time? Because these two kinds of data have completely different meaning. In part two we showed that the value of parameter λ change the structure of CFGMWEM. So we want to show the values of this parameter in modeling. In Table 16 there are maximum likelihood estimation values for parameters in modeling data1 to data 5.

Table 16. Values of Parameter Estimation in Models

Stream and Situation	θ	λ
Wheaton River	0.1015	0.9706
Terme March (high flow)	0.0377	-1
Terme September (low flow)	0.9440	0.7043
Sefaati April (high flow)	0.0385	-1
Sefaati August (low flow)	0.5748	0.4211

We can easily see that when CFGMWEM gains conformity to extreme data, λ parameter takes value between $[-1, 0)$ and when CFGMWEM gains conformity to low flow data, λ parameter takes value between $(0, 1]$. According to test results for Data 1 to Data 5 we suggest that CFGMWEM can be used in every kinds of flows.

We examine that CFGMWEM has best results in Data group 1, Data group 3 and Data group 5. For Data group 2 CFGMWEM has the second best results in modelling. And in modeling Data group 4 the new distribution has the third best results. According to Tables in application part we reach the conclusion that CFGMWEM can be identified as a stream distribution.

Notations

$C_\theta(u, v)$	Copula function
$E(X^k)$	k. th pure moment
$F(x, y)$	two dimensional distribution function
$f(x, y)$	two dimensional probability density function
$F(t)$	one dimensional distribution function
$L(\underline{\theta}; x_1, x_2, \dots, x_n)$	maximum likelihood function
$\log(L(\underline{\theta}; x_1, x_2, \dots, x_n))$	logarithm of maximum likelihood function
$M_X(v)$	moment generating function
M^r	pure sampling moment
$r(t)$	hazard rate function
$S(t)$	survival function
$Var(X)$	variance
$\hat{\cdot}$	parameter estimator
$\Gamma(x)$	gamma function
L_i	L samples
λ_i	L moments

REFERENCES

- [1] Akaike, H. A New Look at the Statistical Model Identification. IEEE Transactions on Automatic Control., (1974). 19(6), 716-723.
- [2] Choulakian, V. and M.A. Stephens. Goodness-of-fit tests for the generalized Pareto distribution. Technometrics., (2001). 478-484.
- [3] Henderson, R. and Diettrich, J. Statistical analysis of river flow data in the Horizons Region: NIWA Client Report: CHC., (2007) 2006-154.
- [4] Merovci, F. and Puka L. Transmuted Pareto distribution. ProbStat Forum., (2014). 1-11.
- [5] Næss S. K. Application of the Kolmogorov-Smirnov test to CMB data: Is the universe really weakly random? Astronomy and Astrophysics (2012). A17, 1-4.
- [6] Nelsen, R.B. An Introduction to Copulas. New York: (2006). Springer.
- [7] Opere, A.O., Mkhundi, S. and Willems, P. At site flood frequency analysis for the Nile equatorial basins. Physics and Chemistry of the Earth., (2006). 31, 919-927.
- [8] Ross, S.M. Introduction to Probability and Statistics for Engineers and Scientists. Third Edition. Elsevier Academic Press: (2004). 504-514.
- [9] Sklar, A. Fonctions de repartition an dimensionset leurs marges. Publ. Inst. Statist. Univ. Paris.; (1959). 229-231.
- [10] Snipes, M. and Taylorn, C.D. Model selection and Akaike Information Criteria: An example from wine ratings and prices. Wine Economics and Policy, (2014). 3(1), 3-9.
- [11] Zaidman, M.D., Keller, V. and Young, A.R. Probability Distributions for “x-day” Daily Mean Flow Events. Bristol: Environment Agency., (2002).