# Modeling hydrologic data by means of re-parametrization of Beta-Singh-Maddala distribution

F. Condino[1,*], F. Domma[1] and S. Franceschi[1]

[1] *Department of Economics, Statistics and Finance "Giovanni Anania", University of Calabria; francesca.condino@unical.it, filippo.domma@unical.it, sara.franceschi@unical.it*
**Corresponding author*

***Abstract.*** *In this paper we propose a new parametrization of the four-parameters Beta-Singh-Maddala distribution suitable for the context of hydrologic studies. With this aim, we reparameterize the Beta-Singh-Maddala distribution to make its parameters directly interpretable in terms of measures much more relevant for their practical use than the classical shape, location and scale parameters of the parametric families generally used for modeling hydrologic events. Moreover, in order to evaluate how climatic or physic characteristics could affect these measures, we will express them as functions of a set of covariates that could have an effect separately and/or simultaneously.*

***Keywords.*** *Regional flood frequency; Extreme events; Regression.*

## 1 Introduction

Nowadays, the occurrence and impact of hydrologic extreme events and their possible relationship with climate change represents a crucial theme for human life. In this context, the statistics of extremes plays a fundamental role and represents a strategic tools for the assessment of current and future exposure to risks. The improvement of models for better exploring observed extremes, with an emphasis on flood quantiles, are strategic activities for the assessment of current and future exposure to risks and the development of some appropriate tools for accurately describing some particular phenomena are crucial. With this aim, Domma and Condino [1] propose the use of two new four-parameters distribution functions, namely the Beta-Dagum and Beta-Singh-Maddala distributions which seem to possess the main suitable features to be used for the analysis of extreme events. Furthermore, following Domma et al. [2], in this paper we consider the reparameterization of the four-parameters Beta-Singh-Maddala (Beta-SM4) distribution in order to make its parameters directly interpretable in terms of median, return level and return period. So, the new reformulation allows of making the parameters of the distribution directly interpretable in terms of measures much more relevant for their practical use than the classical shape, location and scale parameters of the parametric families used as in modeling hydrologic events. Moreover, in order to evaluate how climatic or physic characteristics could affect these measures, we will express them as functions of a set of covariates that could have an effect separately and/or simultaneously.

## 2 Reparameterization of the Beta-SM4 distribution

The Beta-SM4 distribution, in its original parameterization, has the following distribution function (*df*):

$$F_{Beta-SM4}(x;\boldsymbol{\xi}) = \left[1-(1+\gamma_2 x^{\gamma_3})^{-\gamma_1}\right]^a \tag{1}$$

where $\boldsymbol{\xi}' = (\gamma_1,\gamma_2,\gamma_3,a)$, with $a>0$ and $\gamma_i>0$ for $i=1,2,3$. The probability density function (*pdf*) is given by $f_{Beta-SM4}(x;\boldsymbol{\xi}) = a\,[F_{SM}(x;\boldsymbol{\gamma})]^{a-1}\,f_{SM}(x;\boldsymbol{\gamma})$. where $F_{SM}(x;\boldsymbol{\gamma})$ and $f_{SM}(x;\boldsymbol{\gamma})$ are, respectively, the *df* and *pdf* of SM distribution.

Following the proposal of [2], we consider the possibility of reformulate the Beta-SM4$(\gamma_1,\gamma_2,\gamma_3,a)$ in terms of new parameters, $I_j$, $j=1,\ldots,4$, that are indicators describing some peculiarities of hydrologic data distribution and such that there exist a one-to-one transformation of the kind $I_j = g_j(\gamma_1,\gamma_2,\gamma_3,a)$, $j=1,\ldots,4$, in order to have a unique solution in terms of $\gamma_1,\gamma_2,\gamma_3$ and $a$:

$$\begin{cases}\gamma_1 = \gamma_1(I_1,I_2,I_3,I_4)\\ \gamma_2 = \gamma_2(I_1,I_2,I_3,I_4)\\ \gamma_3 = \gamma_3(I_1,I_2,I_3,I_4)\\ a = a(I_1,I_2,I_3,I_4)\end{cases}. \tag{2}$$

Substituting the solution (2) in (1), it is possible to obtain the expressions of the *cdf* in terms of the chosen indicators. Analogously to the generalized linear models, the measures $I_j$ are related to the set of covariates, $\mathbf{x}_{j,i}$, by $I_{j,i} = h_j(\mathbf{x}_{j,i},\boldsymbol{\gamma}_j)$, where $h_j(.)$ are suitable link function.

### 2.1 Formulation in terms of median and return level

In this paper, the original parameters are substituted by the following one-to-one transformation $(\gamma_1,\gamma_2,\gamma_3,a)\mapsto(\tau,me,x_0,a)$ where $\tau=\frac{1}{\gamma_1}$, $me$ is the median of distribution, given by

$$me_{Beta-SM4}(p) = \gamma_2^{-\frac{1}{\gamma_3}}\left[(1-0.5^{\frac{1}{a}})^{-\frac{1}{\gamma_1}}-1\right]^{\frac{1}{\gamma_3}}$$

and $x_0$ is the return level, corresponding to a pre-fixed return period $\pi_{x_0}$, i.e.

$$x_0(\pi_{x_0}) = \gamma_2^{-\frac{1}{\gamma_3}}\left\{\left[1-\left(1-\frac{1}{\pi_{x_0}}\right)^{\frac{1}{a}}\right]^{-\frac{1}{\gamma_1}}-1\right\}^{\frac{1}{\gamma_3}}.$$

After simple algebra, we obtain

$$\begin{cases}\gamma_1 = \frac{1}{\tau}\\ \gamma_2 = [(1-0.5^{1/a})^{-\tau}-1]\cdot me^{-\frac{\log\left\{\left[1-\left(1-\frac{1}{\pi_{x_0}}\right)^{1/a}\right]^{-\tau}-1\right\}-\log\left\{[1-0.5^{1/a}]^{-\tau}-1\right\}}{\log x_0-\log me}}\\ \gamma_3 = \frac{\log\left\{\left[1-\left(1-\frac{1}{\pi_{x_0}}\right)^{1/a}\right]^{-\tau}-1\right\}-\log\left\{[1-0.5^{1/a}]^{-\tau}-1\right\}}{\log x_0-\log me}\\ a = a\end{cases}. \tag{3}$$

It is immediate, from (1), to obtain the new expression of *cdf* of Beta-SM4 *r.v.* in terms of median and return level.

# 3 Application

In order to show the usefulness of the proposed model, we consider the data from Hydroclimatic Data Network of U.S. Geological Survey (USGS). In particular, we focuses our attention on annual peak flows considered in [3] for basins located in Texas Region. We consider the area of drainage (A, in $km^2$), the slope of main channel (S, in $m/km$), the mean elevation of drainage basin over MSL (E, in $m$) and the length of main channel from divide to gauge (L, in $km$), as covariates and a return period of 50 years, as in the cited paper. Therefore, in this example, we investigate the direct effects of the covariates on the median and the 50-years return level, using the reparametrization in (3) and choosing $\exp(.)$ as the log-link function. We consider no covariate effects on the remaining parameters. Table 1 reports maximum likelihood estimates (MLEs) of the parameters, the corresponding standard errors, $t$-tests and $p$-values, related to the four indicators $I_1 = \tau$, $I_2 = me$, $I_3 = x_0$ and $I_4 = a$. As we expected, many of the considered variables seem to have a significant influence on annual peak flows, in particular with regards to its median value and 50-years return level. Finally, Figure 1 shows empirical and fitted distribution obtained inserting the sample mean for covariates in the expressions of $me$ and $x_0$, to simulate the case of a representative basin which well summarizes the Texas Region peak flows data.

| Covariate | Estimate | SE | t | p-value |
|---|---|---|---|---|
| | | $\tau = \exp(\boldsymbol{x}_{1,i}, \gamma_1)$ | | |
| Intercept | -2.118 | 1.412 | -1.500 | 0.1336 |
| | | $me = \exp(\boldsymbol{x}_{2,i}, \gamma_2)$ | | |
| Intercept | 8.252 | $5.260 \times 10^{-2}$ | 156.875 | $< 0.001$ |
| A | $1.833 \times 10^{-5}$ | $3.597 \times 10^{-6}$ | 5.097 | $< 0.001$ |
| E | $-4.111 \times 10^{-4}$ | $3.412 \times 10^{-5}$ | $-12.049$ | $< 0.001$ |
| L | $4.869 \times 10^{-3}$ | $2.040 \times 10^{-4}$ | 23.865 | $< 0.001$ |
| | | $x_0 = \exp(\boldsymbol{x}_{3,i}, \gamma_3)$ | | |
| Intercept | 10.224 | $8.468 \times 10^{-2}$ | 120.737 | $< 0.001$ |
| S | $1.368 \times 10^{-2}$ | $5.655 \times 10^{-3}$ | 2.419 | 0.0156 |
| L | $3.090 \times 10^{-3}$ | $2.690 \times 10^{-4}$ | 11.486 | $< 0.001$ |
| | | $a = \exp(\boldsymbol{x}_{4,i}, \gamma_4)$ | | |
| Intercept | $5.00 \times 10^{-1}$ | $4.347 \times 10^{-1}$ | 1.150 | 0.250 |

Table 1: MLEs of the parameters (log-likelihood: -29103.63)

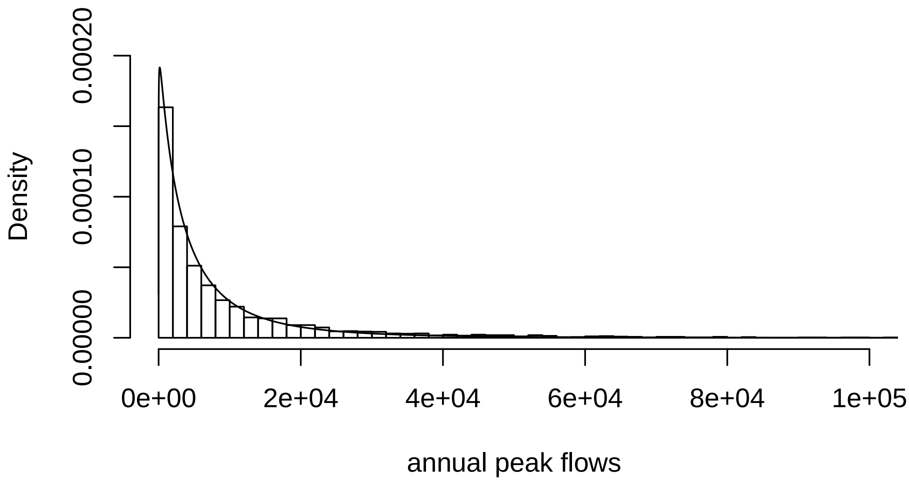


Figure 1: Empirical and fitted Beta-SM4 distribution

## References

[1] Domma, F. and Condino F. (2017) Use of the Beta-Dagum and Beta-Singh-Maddala distributions for modeling hydrologic data. *Stochastic Environmental Research and Risk Assessment* **31**, 799–813.

[2] Domma, F., Condino, F., Giordano S. (2018). A New Formulation of the Dagum Distribution in terms of Income Inequality and Poverty Measures. *Physica A: Statistical Mechanics and its Applications* **511**, 104–126.

[3] Latraverse, M., Rasmussen, P.F., Bobe, B. (2002). Regional estimation of flood quantiles: Parametric versus nonparametric regression models. *Water Resources Research* **38**, 1-1-1-11.