IMPACT OF THE LOCATION PARAMETER ON GAMMA DISTRIBUTION PERFORMANCE IN HYDROLOGICAL APPLICATIONS USING SYNTHETIC SERIES

Maria Luíza Teófilo Gandini [1] & Paulo Ivo Braga de Queiroz [2]

Keywords – Statistical analysis, Moment methods, Extremes events.

INTRODUCTION

The Pearson Type III distribution, also known as the three-parameter Gamma distribution (Gamma 3P), is widely used in hydrology and meteorology to model streamflow and extreme precipitation because it can represent positively skewed data. It consists of the Gamma distribution augmented by a location parameter (γ). Its estimation is well known to be unstable and highly dependent on the estimation method employed. In hydrological applications, this may lead to substantial differences in extreme quantile estimates, affecting risk assessment and hydraulic design. Stacy and Mihram (1965) highlighted the applicability of the Gamma 3P distribution in cases where distributions such as the exponential and Weibull are inadequate. However, they also emphasized that its parameter estimation requires solving equations without closed-form solutions (Stacy, 1962). Subsequent studies proposed more efficient estimators for small and moderate samples (Balakrishnan & Wang, 2000), while analyses of the parameter space geometry showed that its nontrivial structure imposes additional inferential challenges (Chen & Kotz, 2013). Despite its widespread use, it remains unclear whether the location parameter consistently offers practical advantages over the two-parameter Gamma distribution (Gamma 2P), a question that motivates this study.

In this context, this study investigates whether the location parameter of the Gamma 3P distribution provides meaningful additional information or merely increases model flexibility without significant practical benefit. For this purpose, the performance of the Gamma 2P and Gamma 3P distributions is compared using synthetic and observed datasets. The comparison considers goodness-of-fit and quantile estimation via the method of moments. The objective is to assess whether the improvement achieved by the Gamma 3P distribution justifies its greater computational and inferential complexity. Specifically, the study evaluates the influence of the location parameter with respect to two aspects: (i) goodness-of-fit and (ii) estimation of design hydrological quantiles.

MATERIALS AND METHODS

Synthetic datasets were generated in the Python environment for three sample sizes (n = 25, 50, and 100), representing short, intermediate, and long hydrological records. Four classes of theoretical skewness were defined: low (0.4 - 0.8), moderate (0.8 - 1.2), high (1.2 - 2.0), and very high (2.0 - 4.0). These intervals were adopted as descriptive conventions because, for the Gamma 3P distribution, the relationship between skewness and other distributional properties also depends on the location parameter; thus, no universally accepted classification limits exist across all applications (Chen & Kotz, 2013; Veazie, 2023). The Gamma 2P distribution was parameterized by the scale (α) and shape (β) parameters, whereas the Gamma 3P distribution additionally included the location parameter (γ), such that $y = x - \gamma$, with $y \sim \text{Gamma}(\alpha, \beta)$. The adopted probability density function is

1) Universidade Federal de Ouro Preto, Campus Morro do Cruzeiro, Ouro Preto, MG, Brasil, (31) 3559-1541, maria.gandini@ufop.edu.br.
2) Instituto Tecnológico de Aeronáutica, Departamento de Ciência e Tecnologia Aeroespacial, São José dos Campos, SP, Brasil, (12) 3947-6967, pi@ita.br.

given in Equation (1), and the Gamma function in Equation (2):

$$f(x) = \frac{1}{\alpha\Gamma(\beta)}\left(\frac{x-\gamma}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{x-\gamma}{\alpha}\right)\right] \text{ with } x, \alpha \text{ and } \beta > 0 \quad (1)$$

$$\Gamma(\beta) = \int_0^\infty x^{\beta-1}e^{-x}\,dx \quad (2)$$

For each simulation, the skewness coefficient (G) was randomly sampled from a uniform distribution within its respective skewness class, and the shape parameter was calculated as $\beta = (2/G)^2$. The scale parameter (α) was randomly sampled from 1 to 50. Whereas, for the Gamma 3P distribution, the location parameter (γ) was sampled from 0 to 100, for the Gamma 2P distribution, it was fixed at zero. A total of 1,200 synthetic datasets were generated, with 600 for each distribution, equally distributed across sample sizes and skewness classes (50 per class and size). The experimental design included two scenarios: (i) datasets generated from the Gamma 2P distribution and (ii) datasets generated from the Gamma 3P distribution, enabling evaluation under both correctly specified and misspecified conditions.

The parameters α, β, and γ (when applicable) were estimated using the method of moments and compared with the true generating values. This method was selected because it is widely used in hydrological applications, provides analytical solutions for both distributions, and isolates the effect of the location parameter without introducing bias from the estimation method. The estimators are given in Equations 3, 4 and 5, where $\bar{x}$ is the sample mean, σ the sample standard deviation, and G the sample skewness coefficient (Naghettini & Silva, 2017).

$$\text{For Gama 2P: } \alpha = \frac{\sigma^2}{\bar{x}} \qquad \text{For Gama 3P: } \alpha = \frac{\sigma G}{2} \quad (3)$$

$$\text{For Gama 2P: } \beta = \left(\frac{\bar{x}}{\sigma}\right)^2 \qquad \text{For Gama 3P: } \beta = \frac{4}{G^2} \quad (4)$$

$$\text{For Gama 2P: } \gamma = 0 \qquad \text{For Gama 3P: } \gamma = \bar{x} - \frac{2\sigma}{G} \quad (5)$$

Goodness of fit was assessed using the Kolmogorov-Smirnov (KS) test at a 0.10 significance level, following the methodology compiled by Gandini and Queiroz (2024). Empirical cumulative distribution functions were obtained using the Weibull plotting-position formula and compared with theoretical distributions. Quantiles for return periods of 2, 50, and 100 years (frequent, intermediate, and extreme events) were also estimated. Model performance was summarized using KS statistics (D_{max} and p-value) and relative quantile errors. This framework allowed the simultaneous evaluation of goodness of fit and the practical implications for extreme-event estimation. As a complementary validation, the methodology was applied to four observed hydrological datasets of annual maxima (two precipitation and two streamflow series), spanning varying sample sizes and degrees of skewness.

RESULTS AND DISCUSSIONS

The results, summarized in Table 1, were analyzed with respect to two aspects: goodness-of-fit and the ability to estimate design quantiles for different return periods. This framework allowed an assessment of whether the inclusion of the location parameter (γ) in the Gamma 3P provides practical advantages that justify its use in hydrological applications.

Regarding the recovery of original parameter values, when datasets were generated from the Gamma 2P, including the location parameter (γ) provided no benefit, since the data-generating process did not include a location component. Under these conditions, the Gamma 3P distribution introduced an additional parameter without consistent improvements in goodness-of-fit or quantile estimation. Conversely, when the true underlying distribution was Gamma 3P, location-parameter

estimation became more stable as both sample size and skewness increased, indicating that accurate recovery of γ depends on the amount of information available in the sample.

Table 1 – Summary of the results of the synthetic series.

True distribution	Sample size	Skewness	D_{max}		P-value		Rejection rate (%)		Relative error of Tr_2 (%)		Relative error of Tr_{50} (%)		Relative error of Tr_{100} (%)	
			G2P	G3P	G2P	G3P	G2P	G3P	G2P	G3P	G2P	G3P	G2P	G3P
Gamma 2P (G2P)	25	Low	0.13	0.11	0.78	0.80	0	2	0.58	0.29	-0.53	0.36	-0.63	0.58
		Moderate	0.13	0.11	0.70	0.80	2	0	0.54	1.56	-1.34	-1.80	-1.47	-2.07
		High	0.13	0.12	0.72	0.78	0	0	-2.77	2.92	-5.19	-8.95	-5.15	-10.05
		Very high	0.15	0.19	0.62	0.41	0	20	22.72	42.96	-5.21	-8.92	-6.01	-11.28
	50	Low	0.10	0.08	0.72	0.78	0	0	0.07	-0.14	-0.06	-0.14	-0.06	-0.10
		Moderate	0.09	0.09	0.76	0.74	0	0	0.69	0.95	-0.90	-0.77	-1.01	-0.80
		High	0.09	0.09	0.79	0.78	0	2	3.61	6.86	0.18	-1.98	0.00	-2.84
		Very high	0.12	0.22	0.57	0.17	14	62	9.51	27.19	-5.33	-8.61	-5.51	-10.42
	100	Low	0.06	0.06	0.77	0.82	0	0	-0.01	0.35	-1.13	-1.61	-1.23	-1.89
		Moderate	0.06	0.06	0.81	0.77	0	4	-1.51	-1.68	-0.31	-0.21	-0.18	0.13
		High	0.07	0.08	0.71	0.61	2	6	0.94	3.39	-0.61	-2.07	-0.70	-2.47
		Very high	0.08	0.20	0.57	0.13	2	72	11.76	25.60	-3.90	-6.22	-4.34	-7.67
Gamma 3P (G3P)	25	Low	0.12	0.11	0.78	0.83	0	0	0.74	0.13	0.48	2.22	0.32	2.56
		Moderate	0.12	0.11	0.83	0.83	0	0	0.72	-0.31	-6.85	-4.53	-8.01	-4.89
		High	0.13	0.13	0.73	0.79	2	0	5.13	2.59	-3.22	0.01	-5.02	-0.42
		Very high	0.23	0.22	0.21	0.30	38	42	7.79	2.87	-11.69	-5.56	-15.98	-7.33
	50	Low	0.09	0.08	0.76	0.80	0	0	0.56	0.16	-1.33	-0.09	-1.69	-0.05
		Moderate	0.10	0.09	0.71	0.77	0	0	1.38	0.06	-2.24	0.05	-3.06	0.05
		High	0.11	0.09	0.61	0.79	4	0	4.41	2.10	-3.40	-0.62	-5.09	-1.20
		Very high	0.23	0.21	0.05	0.22	86	54	8.97	2.90	-11.15	-4.04	-15.73	-5.61
	100	Low	0.07	0.06	0.75	0.81	0	0	-0.47	-0.89	-2.32	-1.65	-2.68	-1.74
		Moderate	0.07	0.06	0.68	0.76	0	2	1.80	0.58	-3.25	-1.33	-4.25	-1.59
		High	0.11	0.08	0.30	0.60	36	10	4.42	0.52	-4.17	0.79	-6.06	0.91
		Very high	0.24	0.20	0.00	0.10	100	80	8.99	2.82	-11.80	-4.60	-16.33	-5.98

With respect to goodness-of-fit, both distributions exhibited similar overall performance. When the true distribution was Gamma 2P, the mean values of the maximum KS test statistic (D_{max}) and the corresponding p-values were nearly identical, indicating that the additional parameter contributed little to the overall fit. Even when the true distribution was Gamma 3P, differences in goodness-of-fit statistics remained small, suggesting that inclusion of γ provides only limited benefits when performance is assessed solely through global fit criteria. KS test rejection rates confirmed this pattern. For samples of 25 observations, virtually no differences were observed between models. As sample size increased, particularly for samples of 100 observations with high skewness, differences became more noticeable, reflecting the test's greater sensitivity, although still smaller than those observed for quantile estimation.

The analysis of quantiles for return periods of 2, 50, and 100 years showed that when the true distribution was Gamma 2P, estimation errors from both models were comparable, indicating no substantial advantage of the Gamma 3P distribution for design quantiles. In contrast, when data were generated from Gamma 3P, differences increased with both skewness and sample size, particularly for longer return periods. Under these conditions, Gamma 3P yielded lower quantile errors, especially for the 100-year return period, highlighting the importance of the location parameter for representing extreme hydrological events.

Finally, the improvement provided by the Gamma 3P distribution in estimating extreme quantiles was not always reflected in goodness-of-fit statistics. In several cases, both distributions yielded similar D_{max} values and p-values, whereas more pronounced differences were observed only in quantile estimation. This demonstrates that good overall fit does not necessarily imply improved extreme-event estimation, emphasizing the need to jointly evaluate goodness-of-fit and performance metrics relevant to hydrological design objectives.

The results for the observed datasets are summarized in Table 2. For all datasets, both Gamma 2P and Gamma 3P provided satisfactory fits according to the KS test, with no meaningful differences in D_{max} or p-values. Comparison of design quantiles was limited to return periods consistent with record length, preventing evaluation of longer periods in some cases. Nevertheless, the results

observed were consistent with synthetic data, indicating that the main advantages of Gamma 3P are concentrated in highly skewed series, whereas for low to moderate skewness, both distributions perform similarly.

Table 2 – Summary of the results of the observed series.

Hidroweb code	Series type	Sample size	Skewness	D_{max}		P-value		KS test result		Relative error of Tr_2 (%)		Relative error of Tr_{50} (%)	
				G2P	G3P	G2P	G3P	G2P	G3P	G2P	G3P	G2P	G3P
2043011	Precipitation	80	Low (G = 0.5)	0.05	0.04	0.99	1.00	Non-rejected	Non-rejected	-1.09	-0.80	-0.59	-1.09
2042018	Precipitation	46	Medium (G = 0.8)	0.09	0.07	0.87	0.96	Non-rejected	Non-rejected	3.83	2.64	-	-
40579995	Streamflow	37	High (G = 1.8)	0.11	0.12	0.76	0.63	Non-rejected	Non-rejected	0.31	-3.50	-	-
41180000	Streamflow	66	Very high (G = 2.4)	0.09	0.13	0.64	0.22	Non-rejected	Non-rejected	-0.07	-10.58	-27.50	-22.12

CONCLUSIONS

The results show that the location parameter (γ) offers no systematic advantage of the Gamma 3P distribution across hydrological applications. Its benefits depend on the dataset, with Gamma 2P and Gamma 3P performing similarly in most cases, especially for Gamma 2P datasets and small samples. When the true distribution was Gamma 3P, estimating the location parameter was more useful for highly skewed datasets, particularly for longer return-period quantiles. These differences were more evident in design quantiles than in goodness-of-fit tests, indicating that goodness-of-fit alone is insufficient for model selection. Results from the observed datasets were consistent with those from the synthetic datasets, reinforcing the conclusion that the location parameter primarily benefits highly skewed datasets.

A limitation of the present study is the use of synthetic datasets and the exclusive adoption of the method of moments for parameter estimation. Although the methodology was also applied to four observed hydrological datasets, this sample is not sufficiently large to support broader generalizations. Future research should therefore evaluate a statistically representative set of observed precipitation and streamflow records and compare different parameter estimation methods to better assess the practical generalizability of the findings reported herein.

REFERENCES

BALAKRISHNAN, N.; WANG, J. (2000). Simple efficient estimation for the three-parameter gamma distribution. *"J. Stat. Plan. Inference"*, 85, 115-126.

CHEN, W. W. S.; KOTZ, S. (2013). The Riemannian Structure of the Three-Parameter Gamma Distribution. *"J. Appl. Math."*, 4, 514-522.

GANDINI, M. L. T.; QUEIROZ, P. I. B. (2024). *"Compilation of methodology for the application of the Kolmogorov-Smirnov test and calculation of the p-value without the use of tables"* in Proceedings of 9th International Symposium on Integrated Water Resources Management (IWRM) | 14th International Workshop on Statistical Hydrology (STAHY) | I EBHE - Encontro Brasileiro de Hidrologia Estatística, Florianópolis-SC, Nov. 2024.

NAGHETTINI, M.; SILVA, A. T. (2017). *"Continuous random variables: probability distributions and their applications in hydrology"*. In: NAGHETTINI, M. (ed) Fundamentals of statistical hydrology. Springer International Publishing AG, Gewerbestrasse, p 660

STACY, E. W. (1962). A generalization of the gamma distribution. *"Annals of Mathematical Statistics"*; 33: 1187-1192.

STACY, E. W.; MIHRAM, G. A. (1965). Parameter estimation for a generalized gamma distribution. *"Technometrics"*; 7: 349-358.

VEAZIE, P; INTRATOR, O; KINOSIAN, B; PHIBBS, C. (2023). Better performance for right-skewed data using an alternative gamma model. *"BMC Medical Research Methodology"*. Dec 1st.