ISSN 2359-1919 (Online)
SCIENTIFIC ARTICLE

Performance comparison of cross-validation techniques in fitting Generalized Extreme Value (GEV) models to monthly low-flow extremes

Comparação do desempenho de técnicas de validação cruzada no ajuste de modelos GEV para extremos de vazões mínimas mensais

Luiz Antônio de Oliveira[1]

[1]Universidade Federal de Uberlândia, Uberlândia, MG, Brasil. E-mail: ao.luiz@gmail.com

How to cite: Oliveira, L. A. (2025). Performance comparison of cross-validation techniques in fitting Generalized Extreme Value (GEV) models to monthly low-flow extremes. *Revista de Gestão de Água da América Latina*, *22*, e21. https://doi.org/10.21168/rega.v22e21

ABSTRACT: This study compares the performance of the K-Fold (K=10) and Leave-One-Out (LOOCV) cross-validation techniques in assessing the goodness-of-fit of Extreme Value Distributions (EVDs) for low-flow hydrological extremes. The analysis employed the Generalized Extreme Value (GEV) distribution and its subfamilies (Gumbel, Weibull, and Fréchet), using daily streamflow data from the Porto Velho gauging station, located in the municipality of Porto Velho, Rondônia state, within the Brazilian Amazon, over the period 1970–2024. A dynamic monthly 10th percentile was adopted as the threshold to define low-flow extremes. Both LOOCV and Stratified K-Fold procedures were applied, with model adequacy evaluated based on the Kolmogorov–Smirnov (KS) statistic. Results indicate that the The LOOCV method consistently exhibits more stable KS statistics across all months and distributions. This pattern indicates a more uniform model performance, with lower sensitivity to specific months or to the presence of extreme events. In contrast, the K-Fold method shows lower KS values but with pronounced fluctuations observed in certain months. These peaks are notably associated with periods of higher exceedance frequencies or the occurrence of minimum extreme streamflows. This behavior reflects the K-Fold method's sensitivity to hydrological extremes: when extreme events are concentrated within specific folds, the KS statistic tends to decrease.

Keywords: Low-Flow Hydrological Extremes; Cross-Validation; Generalized Extreme Value (GEV); Porto Velho; Brazil.

RESUMO: Este estudo compara o desempenho das técnicas de validação cruzada K-Fold (K=10) e Leave-One-Out (LOOCV) na avaliação do ajuste de distribuições de valores extremos (EVDs) aplicadas aos extremos hidrológicos de vazões mínimas. A análise utilizou a distribuição Generalizada de Valores Extremos (GEV) e suas subfamílias (Gumbel, Weibull e Fréchet), empregando dados diários de vazão da estação fluviométrica de Porto Velho, localizada no município de Porto Velho, estado de Rondônia, Amazônia Brasileira, período de 1970 a 2024. Foi adotado como limiar para definição dos extremos de vazões mínimas o percentil 10 dinâmico mensal. Ambos os procedimentos, LOOCV e K-Fold estratificado, foram aplicados, sendo a adequação dos modelos avaliada com base no estatístico Kolmogorov–Smirnov (KS). Os resultados indicam que o método LOOCV apresenta, de forma consistente, valores mais estáveis da estatística KS ao longo de todos os meses e distribuições. Esse padrão sugere um desempenho mais uniforme do modelo, com menor sensibilidade a meses específicos ou à presença de eventos extremos. Em contraste, o método K-Fold apresenta valores mais baixos da estatística KS, porém com flutuações acentuadas observadas em determinados meses. Esses picos estão notavelmente associados a períodos com maiores frequências de excedência ou à ocorrência de vazões mínimas extremas.

Palavras-chave: Vazões Extremas Mínimas; Cross-Validação; Distribuição Generalizada de Valores Extremos (GEV); Porto Velho; Brasil.

INTRODUCTION

The analysis of extreme events is a fundamental pillar of the hydrological and environmental sciences since floods and droughts have a significant impact on water resource management, water security, agriculture, energy, and ecosystems. However, much of the literature focuses on modeling the upper extremes (upper tail), which are associated with floods and maximum precipitation.

Received: June 06, 2025. Revised: September 29, 2025. Accepted: October 12, 2025.

However, analysis of the lower tail of the distribution is equally critical for understanding scarcity events, such as minimum flows, critical reservoir levels, and severe hydrological droughts (Coles, 2001; Beirlant et al., 2004).Threshold definition is a critical step in the modeling process and there are several approaches to choosing it, including graphical methods, statistical criteria based on information criteria, and percentage approaches, which are particularly useful in practical applications in hydrology (Coles, 2001; Beirlant et al., 2004). The choice of the 10% percentile as the lower threshold is advocated in situations where it is necessary to ensure statistical robustness without compromising the representativeness of the extreme events of the lower tail (Hosking & Wallis, 1997).

Statistical modeling of these events often uses extreme value distributions, particularly the generalized extreme value (GEV) distribution and its particulares distributions: Gumbel, Weibull and Fréchet. These distributions are adapted for analyzing the lower tail and can adequately describe the behavior of annual or monthly minimums in hydrological series (Katz et al., 2002; Gumbel, 1958). According to the extreme value theorem, this is the limit distribution of the normalized maxima or minima of a sequence of independent and identically distributed random variables (NASA GMAO, 2025).

In a hydrological context, monthly cross-validation enables us to evaluate the suitability of the GEV distribution for each month of the year, considering seasonal nuances.Cross-validation is a resampling method used to evaluate the predictive performance of statistical models. Cross-validation is a robust methodology for assessing the performance of these distribution models, as it allows us to estimate the model's predictive capacity and stability in the face of data variability (Arlot & Celisse, 2010).

Accordingly, this study presents a performance comparison between two cross-validation techniques: Leave-One-Out Cross-Validation (LOOCV) and Stratified K-Fold Cross-Validation (K=10). The Kolmogorov–Smirnov (KS) statistic served as the primary metric to evaluate model fit, with lower values indicating better alignment between observed and modeled distributions. The daily streamflow records from the Porto Velho gauging station, located in the municipality of Porto Velho, Rondônia state, in the Brazilian Amazon, were employed as a case study to illustrate the application.

METHODS

We used daily flow data recorded at the Porto Velho conventional fluviometric station (code 15400000, National Water Agency – ANA), which is located in the municipality of Porto Velho, state of Rondônia, Brazilian Amazon, covering the period from 1970 to 2024. This data had been organized, filtered and processed in an Excel spreadsheet (Microsoft Office 365). The statistics were processed, where, for each month, the threshold value and exceedances, as well as K-fold and leave-one-out cross-validations, the mean and standard deviation of the KS test, and the adjusted GEV parameters (shape, location and scale).

In the paper, the analysis of extreme low values (lower tail) is performed using a threshold based on the 10th percentile of the data distribution. Specifically, for each month, the script selects all streamflow values below the 10th percentile. This approach aligns with the principles of the GEV method, which is widely used in extreme value theory to focus on the most significant rare events. To fit classical extreme value distributions, which are typically formulated for máxima, the selected low values are reflected by multiplying by −1 (i.e., for each value X, compute Y = -X). This transformation converts the problem of modeling minima into a standard maxima problem, allowing the use of the Generalized Extreme Value (GEV), Gumbel, Weibull, and Fréchet distributions, which are conventionally parameterized for upper tails (Coles, 2001; Katz et al., 2002; Scarrott & MacDonald, 2012).

The methodological approach adopted was based on validating distributions belonging to the Extreme Value Theory (EVT) family, which provides a mathematical basis for modeling the behavior of distribution tails adequately (Embrechts et al., 1997; Katz et al., 2002; MathWorks, 2025). In the context of this work, parametric modeling was based on the generalized extreme value (GEV) distribution and its particulares distributions: Gumbel, Weibull and Fréchet. These were adapted for the analysis of the lower tail, for which the cumulative distribution function (CDF), (Hosking & Wallis, 1987), is given by:

$$F(x) - \exp\{-[1+\ \xi\left(\frac{x-\mu}{\sigma}\right)]^{\ 1/\xi}\}$$ Equation 1

To: + $1+\xi\left(\frac{x-\mu}{\sigma}\right)>0$:

- μ - location parameter;
- σ - scale parameter;
- ξ - shape parameter.

The parameter ξ defines the type of tail of the distribution (Katz et al., 2002):

- ξ < 0 - limited tail (Weibull-type);
- ξ = 0 - exponential tail (Gumbel-type);
- ξ > 0 - heavy tail (Fréchet-type).

The best fit for each of the GEV distribution types, minimum extremes and 10% percentile thresholds was validated by comparing the stratified K-fold and Leave-One-Out cross-validation (LOOCV) methods. This approach was used to assess how well the model matched the observations representing the lower tail of the distribution, i.e. the most severe events with low values.

Cross-validation is a fundamental tool for assessing the generalizability of statistical models, particularly in the context of extreme value analysis. Is especially important due to the scarcity and high leverage of extreme events. The fundamental idea is to partition the data into a set of training samples, used to fit the model, and a set of validation samples, used to test the model's predictive accuracy. This process is repeated several times, and the results are aggregated to provide a robust estimate of model performance (Stone, 1974; Renard et al., 2013).

In the LOOCV procedure, one observation (xi) from the sample of extremes is left out in each iteration. The GEV model is then fitted to the remaining N-1 observations. Using the obtained parameters, the GEV cumulative distribution function (CDF) was calculated for the excluded observation, and this theoretical CDF was then compared with the empirical CDF constructed iteratively throughout the cross-validation process. The maximum distance between the empirical and theoretical CDFs of the GEV, adjusted for each iteration, was evaluated using the Kolmogorov–Smirnov (KS) test adapted to the context of the left-out data. This process was repeated for each of the N observations in the sample of extremes below the threshold. At the end of the iterations, the following indicators were calculated to assess the robustness of the model: the mean and standard deviation of the KS statistic values obtained in all iterations and the proportion of iterations in which the null hypothesis of the KS test (that the data follow the adjusted GEV) was not rejected at the adopted significance level (α = 5%).

Stratified k-fold cross-validation is a statistical technique used to evaluate predictive models. It seeks to estimate how well the model can generalize to unseen data. The technique involves dividing the data set into K subsets (or 'folds') of roughly equal size. This process is then repeated K times, with K-1 folds being used to train the model in each iteration and one fold being reserved for testing (validation). After K iterations have been completed, the average performance indicator is calculated. In this study, the data set was stratified to obtain ten folds with a proportional distribution of hydrological minimum extremes. In each iteration, nine folds are used to fit the model, and one separate fold is used for testing. Then, the distribution of extreme values (GEV, Gumbel, Weibull or Fréchet) is fitted to the training data (i.e. nine folds), and the CDF of the model is obtained with the parameters fitted to the training data. The empirical CDF of the test fold data is calculated, consisting of the cumulative proportion of the ordered values. The Kolmogorov–Smirnov (KS) test is used to measure the quality of the model's fit. This test calculates the maximum absolute difference between the empirical and theoretical CDFs (Massey Junior, 1951). This process is repeated until each fold has been used for testing. Finally, the mean and standard deviation of the KS values obtained in the ten iterations were calculated.

To evaluate the fit of the GEV distribution, a method was employed that combined the Kolmogorov–Smirnov (KS) test with stratified k-fold and leave-one-out cross-validation. This mitigated the bias inherent in using the KS test on parameters estimated from the data itself. The data set was divided into subsets of approximately equal size. In each iteration, the GEV model was fitted using k-1 subsets (the training set) and the fit was evaluated using the remaining subset (the test set). The KS statistic was then calculated by comparing the empirical distribution function of the test set with the fitted GEV distribution function of the training set. This process was repeated for each subset and the KS statistics were summarized in terms of mean and standard deviation. This provided a robust assessment of the model's adherence with an emphasis on its ability to generalize beyond the sample.

The uncertainty method used are stratified K-fold and Leave-One-Out cross-validation (LOOCV) to measure the variability and uncertainty of the model's performance. The final metrics that express this uncertainty are the Standard Deviation and, more specifically, the Standard Error of the Mean (SEM) of the Kolmogorov-Smirnov scores obtained from the cross-validation folds.

RESULTS

Tables 1 and 2 summarize the performance of the models evaluated by the average KS value obtained in each round of cross-validation (KS_ Mean) and uncertainty metrics (Standard Deviation - STD and Standard Error of the Mean - SEM), comparing the Leave-One-Out (LOOC) method with the k-fold stratified cross-validation method (k = 10) in that order.

The distribution with the lowest KS_CV_Mean (best average fit) and, ideally, with low KS_CV_Std and KS_CV_SEM (most stable and reliable result).

Table 1. Summary results of the Leave-One-Out (LOOC) cross-validation adjustments.

Mes	KS_Mean GEV	KS_Std GEV	KS_SEM GEV	KS_Mean Weibull	KS_Std Weibull	KS_SEM Weibull	KS_Mean Gumbel	KS_Std Gumbel	KS_SEM Gumbel	KS_Mean Frechet	KS_Std Frechet	KS_SEM Frechet
1	0.549	0.182	0.040	0.546	0.205	0.045	0.574	0.197	0.043	0.574	0.197	0.043
2	0.676	0.136	0.036	0.498	0.178	0.048	0.513	0.201	0.054	0.846	0.181	0.048
3	0.643	0.101	0.027	0.540	0.274	0.073	0.554	0.274	0.073	0.843	0.146	0.039
4	0.660	0.065	0.020	0.405	0.134	0.040	0.419	0.142	0.043	0.448	0.190	0.057
5	0.706	0.127	0.031	0.550	0.226	0.055	0.563	0.228	0.055	0.578	0.238	0.058
6	0.612	0.090	0.022	0.530	0.183	0.044	0.545	0.216	0.052	0.545	0.216	0.053
7	0.609	0.220	0.053	0.489	0.269	0.065	0.488	0.274	0.066	0.490	0.276	0.067
8	0.448	0.188	0.047	0.466	0.179	0.045	0.491	0.201	0.050	0.492	0.201	0.050
9	0.660	0.216	0.058	0.663	0.214	0.057	0.680	0.216	0.058	0.696	0.221	0.059
10	0.706	0.173	0.048	0.534	0.254	0.070	0.542	0.255	0.071	0.543	0.257	0.071
11	0.615	0.211	0.050	0.582	0.213	0.050	0.597	0.223	0.052	0.606	0.226	0.053
12	0.695	0.068	0.016	0.534	0.212	0.049	0.548	0.229	0.053	0.861	0.092	0.021

Source: author.

In the analysis of Table 1, the LOOCV method showed that the KS_CV_Mean values varied moderately between the distributions and between the different runs. The Weibull (from 0.405 to 0.663), Gumbel (from 0.419 to 0.680) and Fréchet (from 0.490 to 0.861) distributions exhibited slightly lower values than the GEV distribution (from 0,448 to 0,706). The Gumbel and Fréchet distributions fluctuated depending on the sample, with the Fréchet distribution showing high values in some cases (up to 0.861), which could indicate sensitivity to outliers.

Dominance of the Weibull Distribution: The Weibull distribution is the clear winner, being selected as the best in 9 out of the 12 months. This suggests that for low-flow events (lower extremes) at this station, the Weibull distribution generally provides the best-fitting model. GEV and Gumbel as Alternatives: The GEV was the best in 2 months (August, September) and Gumbel in 1 month (July). They appear as secondary alternatives, mainly during the transition months into the dry season. The GEV distribution consistently exhibits the highest stability (lowest Standard Deviation and Standard Error of the Mean) in almost every month. This means the GEV's performance is more predictable and less sensitive to the removal of a specific year from the data.

The standard deviation values are generally quite high (often above 0.200), especially for the Weibull distribution. This indicates that the extreme value modeling for this hydrological station has considerable uncertainty. The results can change significantly depending on the years included in the analysis, which reinforces the importance of using cross-validation to expose this sensitivity.

Table 2. Summary results of the k-fold stratified cross-validation adjustments (k = 10).

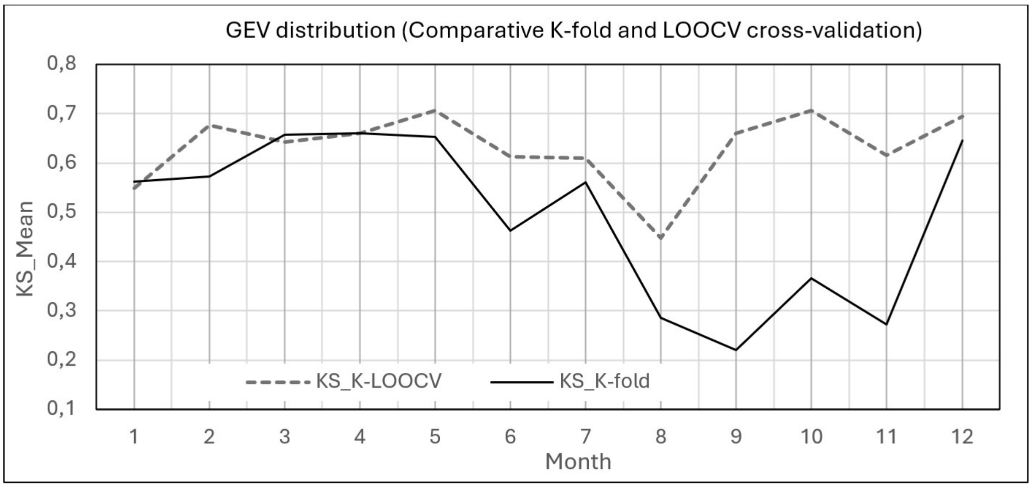
Mes	KS_Mean GEV	KS_Std GEV	KS_SEM GEV	KS_Mean Weibull	KS_Std Weibull	KS_SEM Weibull	KS_Mean Gumbel	KS_Std Gumbel	KS_SEM Gumbel	KS_Mean Frechet	KS_Std Frechet	KS_SEM Frechet
1	0.562	0.216	0.068	0.239	0.083	0.026	0.264	0.082	0.026	0.264	0.082	0.026
2	0.573	0.164	0.052	0.222	0.083	0.026	0.247	0.064	0.020	0.851	0.035	0.011
3	0.658	0.029	0.009	0.204	0.070	0.022	0.216	0.082	0.026	0.853	0.038	0.012
4	0.660	0.055	0.018	0.204	0.059	0.019	0.211	0.058	0.018	0.211	0.058	0.018
5	0.653	0.041	0.013	0.183	0.057	0.018	0.205	0.055	0.017	0.205	0.054	0.017
6	0.463	0.207	0.065	0.263	0.096	0.030	0.279	0.109	0.034	0.279	0.109	0.035
7	0.561	0.242	0.076	0.201	0.052	0.016	0.203	0.048	0.015	0.204	0.050	0.016
8	0.286	0.131	0.041	0.233	0.046	0.015	0.246	0.066	0.021	0.246	0.066	0.021
9	0.220	0.055	0.017	0.202	0.051	0.016	0.228	0.046	0.015	0.227	0.046	0.015
10	0.366	0.126	0.040	0.182	0.064	0.020	0.189	0.059	0.019	0.190	0.059	0.019
11	0.272	0.139	0.044	0.225	0.049	0.016	0.231	0.041	0.013	0.231	0.042	0.013
12	0.645	0.048	0.015	0.194	0.068	0.021	0.196	0.072	0.023	0.821	0.053	0.017

Source: author.

As can be seen in Table 2, the K-fold method significantly reduces the KS_CV_Mean values for all distributions: Weibull (from 0.182 to 0.239), Gumbel (from 0.189 to 0.279), Fréchet (from 0.190 to 0.853) and GEV (from 0.220 to 0.660). Additionally, the Fréchet distribution, which exhibited significant variability in LOOCV, exhibited more consistent values in the k-fold method, indicating the robustness of the stratified validation process.The Weibull distribution is dominant in this validation scenario. It not only shows the best mean performance in the vast majority of months, but its stability (measured by the standard deviation) is consistently good, often being the most stable or very close to the most stable among its direct competitors.

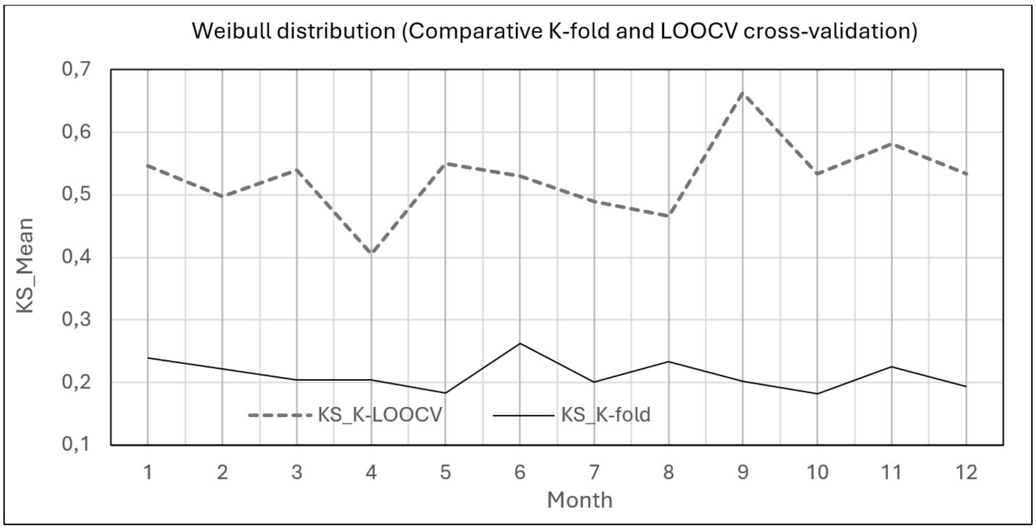
Comparative analysis of the 10-Fold and Leave-One-Year-Out Cross-Validation (LOOCV)

Graphs 1, 2, 3 and 4 shows the behavior of the adjustment values according to the LOOCV or K-fold cross-validation method.



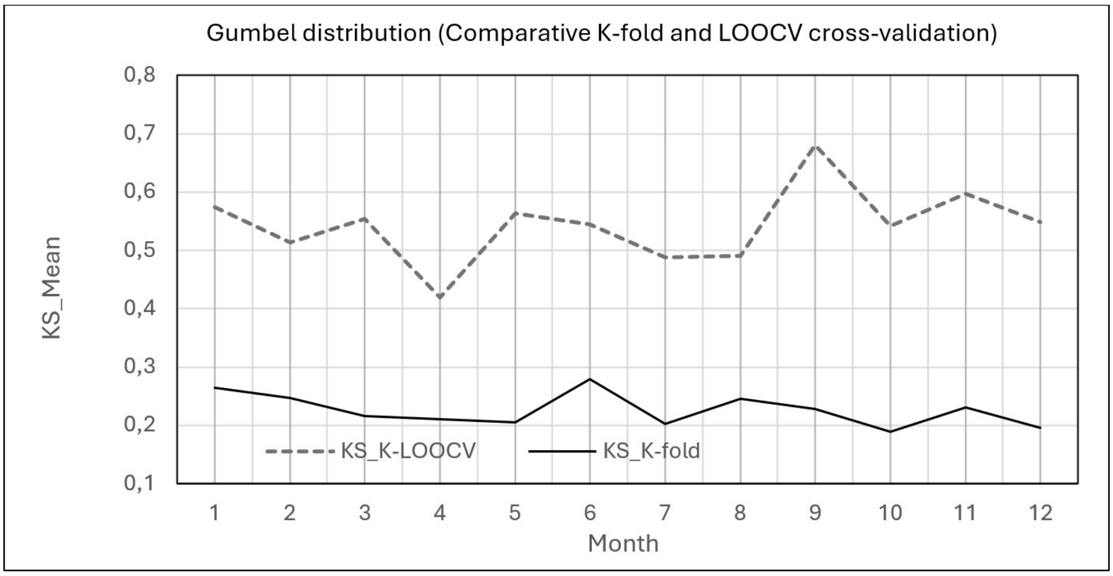
Graph 1. GEV distribution (Comparative K-fold and LOOCV cross-validation)
Source: author.

For the GEV distribution, K-fold's results are more responsive to the underlying variability in the data, especially in months characterized by extreme events—most notably in September, October, and November. This heightened sensitivity suggests that K-fold better captures the underlying structure and seasonality of the data during periods of hydrological extremes.



Graph 2. Weibull distribution (Comparative K-fold and LOOCV cross-validation)
Source: author.

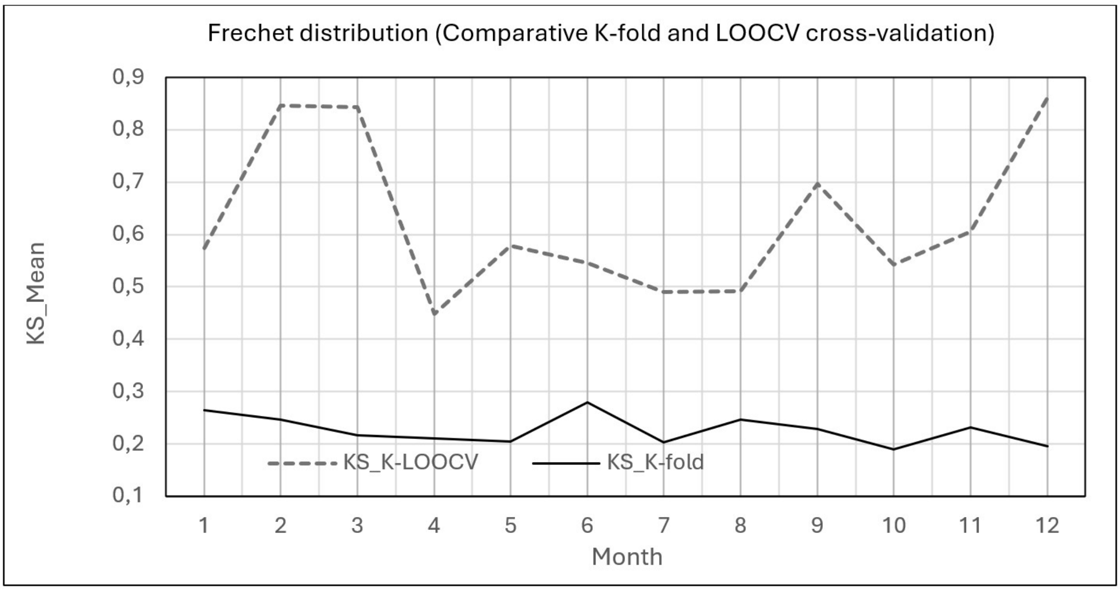
The Weibull distribution, when validated with K-fold cross-validation, achieves the best overall performance among the distributions analyzed, as evidenced by both the low and stable KS statistics. This result underscores the Weibull distribution's suitability for modeling monthly minimum streamflow extremes in this data set, combining accuracy with robustness across different months.



Graph 3. Gumbel distribution (Comparative K-fold and LOOCV cross-validation)

Source: author.

For the Gumbel distribution, the use of K-fold cross-validation proves advantageous, as it presents a better fit with lower KS values. This indicates that K-fold not only mitigates the influence of individual data points but also enhances the accuracy and reliability of the model's performance assessment for this distribution.



Graph 4. Frechet distribution (Comparative K-fold and LOOCV cross-validation)

Source: author.

For the Frechet distribution, K-fold's results are notably more responsive to the underlying variability in the data, especially during months characterized by extreme events. This is particularly evident in March, April, May, June, July, August, September, October, and November, where pronounced peaks in the KS statistic are observed. These fluctuations suggest that K-fold is sensitive to the occurrence of hydrological extremes during these periods, enabling it to better capture the underlying structure and seasonality present in the streamflow data.

A comparative analysis of the KFold and Leave-One-Year-Out Cross-Validation (LOOCV) methodologies reveals critical differences in their assessment of model performance and uncertainty. While the K10-Fold method produced numerically superior metrics, lower mean Kolmogorov-Smirnov (KS) statistics (≈0.20) and significantly lower standard deviations, the LOOCV method is concluded to be the more appropriate and robust validation strategy for this hydrological time-series analysis.

The apparent superiority of the KFold method can be attributed to its random shuffling of the data set, which breaks the inherent temporal structure of the data. By mixing observations from all years within the training and testing folds, this method masks the significant inter-annual variability present in the hydrological record. This leads to an optimistically biased assessment, suggesting a higher degree of model stability and predictive accuracy than is realistic.

In contrast, the models likely perform better in the core of the dry season (e.g., June, July, August) because the river flow is more stable and consistently low. This homogeneity makes it easier for a single probability distribution to capture the behavior of the extremes.

Figures 1 through 12 illustrate the performance of the monthly flow threshold values (m^3/s), from January to December, scaled by decades, with red dashed lines representing the thresholds. These graphs also highlight the outliers, the P10% threshold, and the lower tail of the distribution.

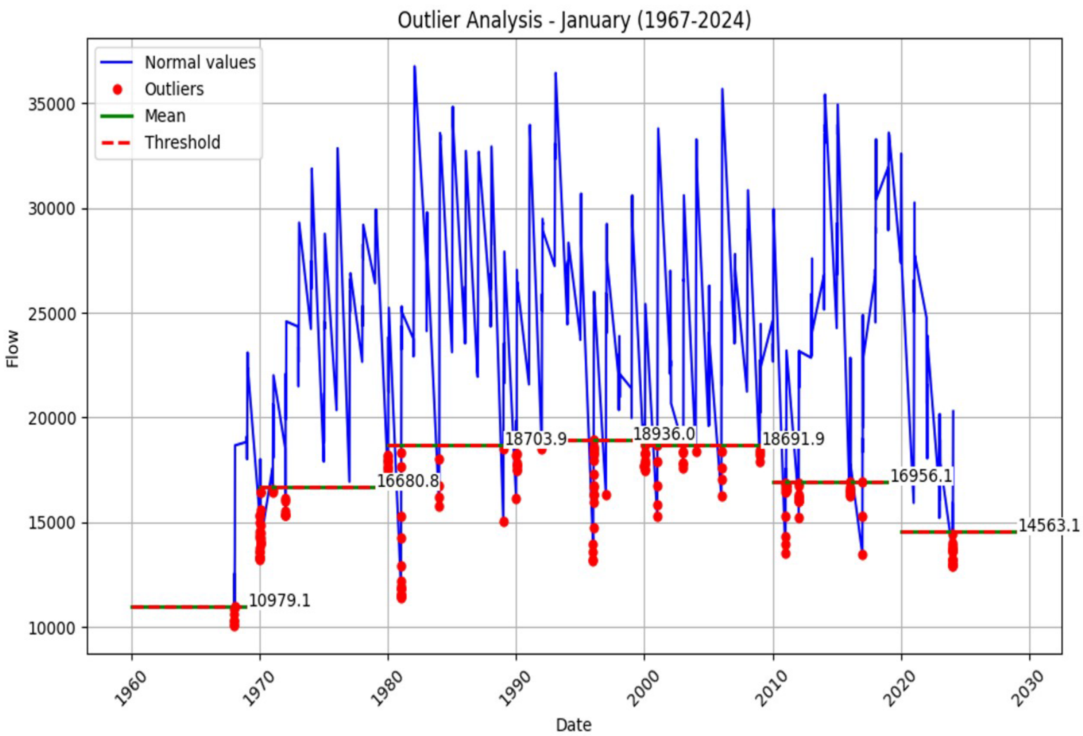


Figure 1. Dynamic flow thresholds (m^3/s) for January, with the P10% threshold, scaled by month and decade.

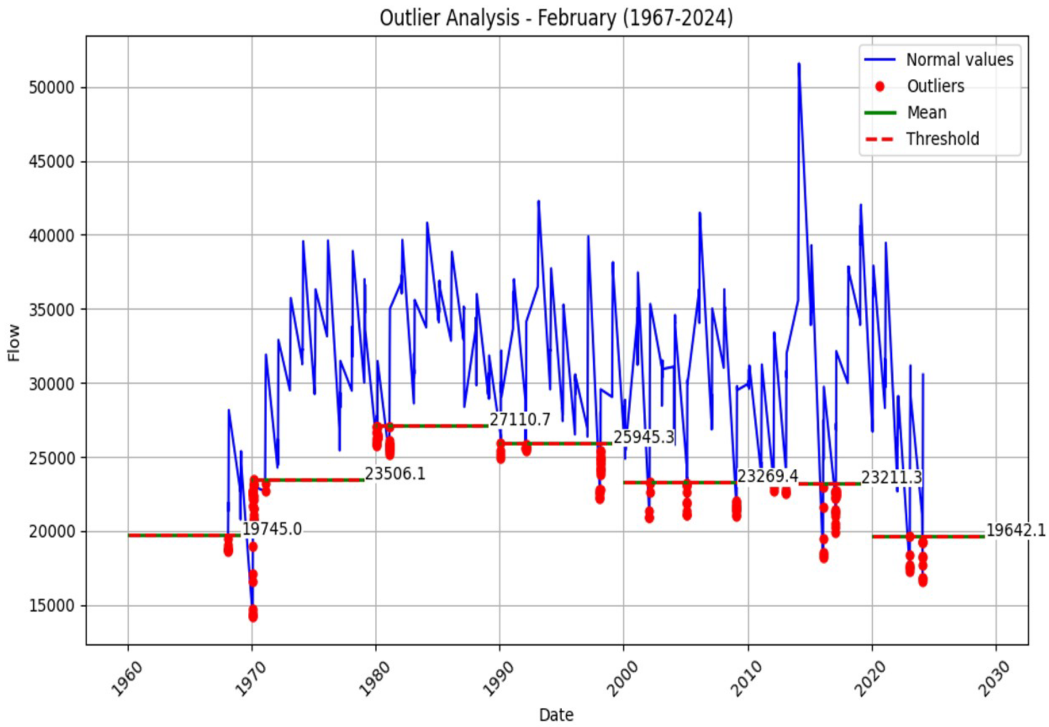


Figure 2. Dynamic flow thresholds (m³/s) for february, with the P10% threshold, scaled by month and decade.

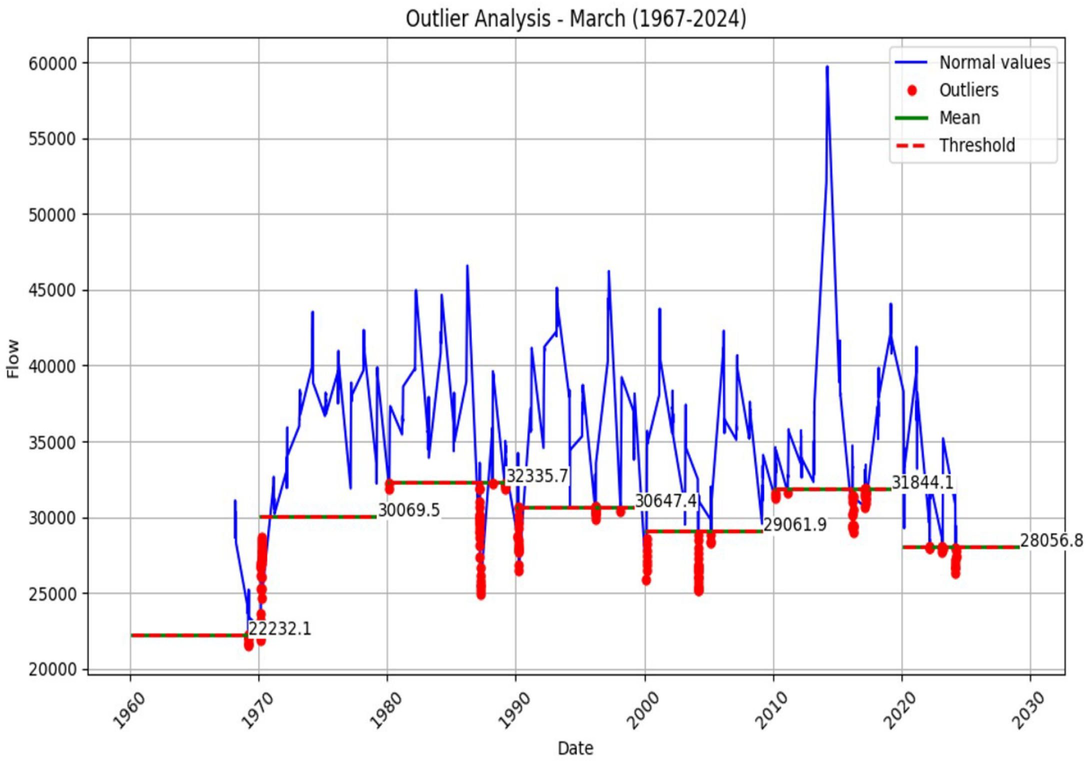


Figure 3. Dynamic flow thresholds (m³/s) for march, with the P10% threshold, scaled by month and decade

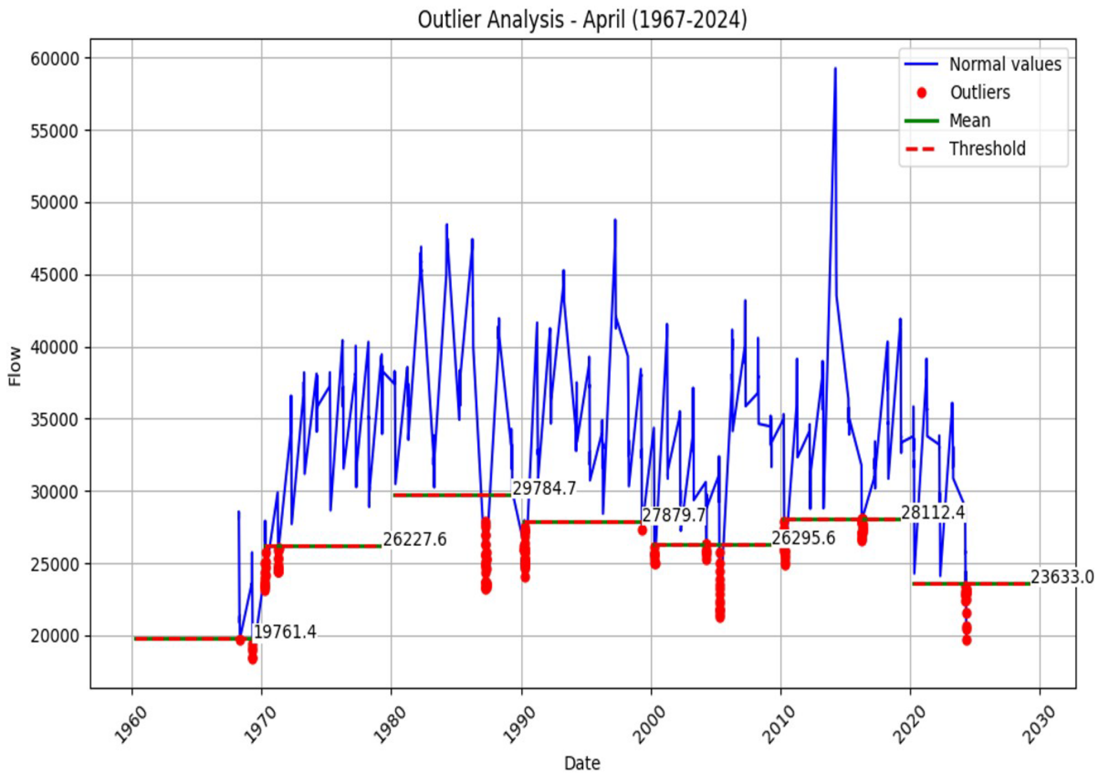


Figure 4. Dynamic flow thresholds (m³/s) for april, with the P10% threshold, scaled by month and decade

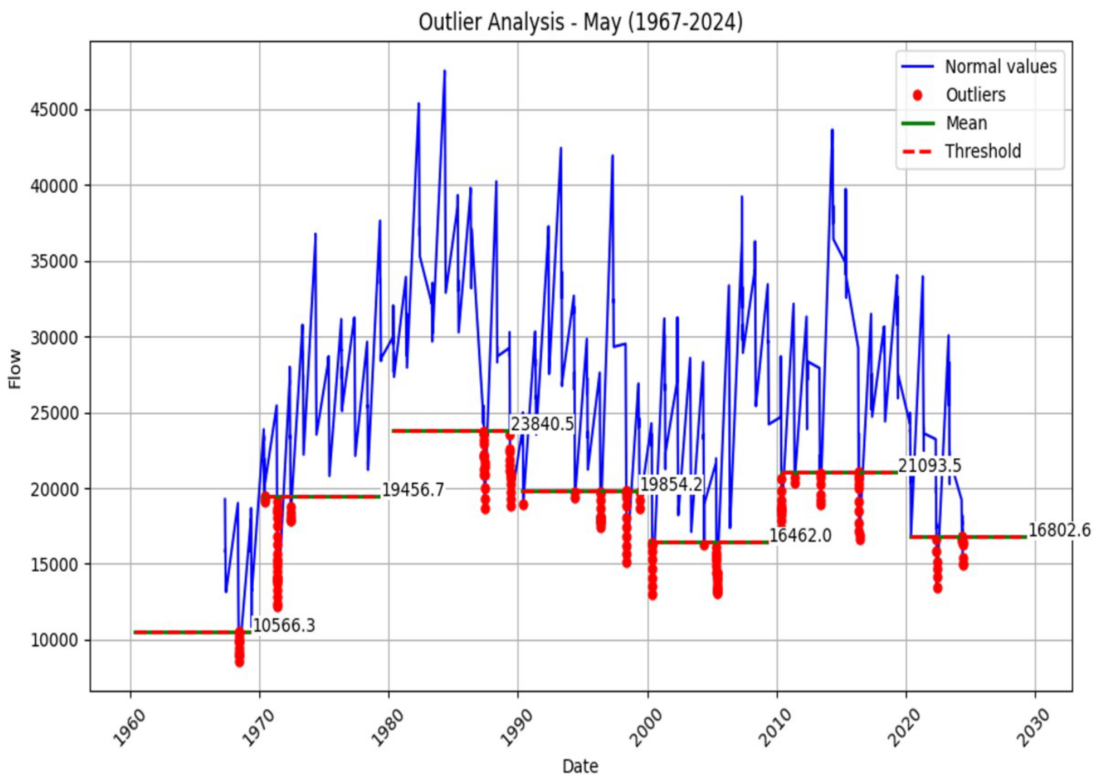


Figure 5. Dynamic flow thresholds (m³/s) for may, with the P10% threshold, scaled by month and decade

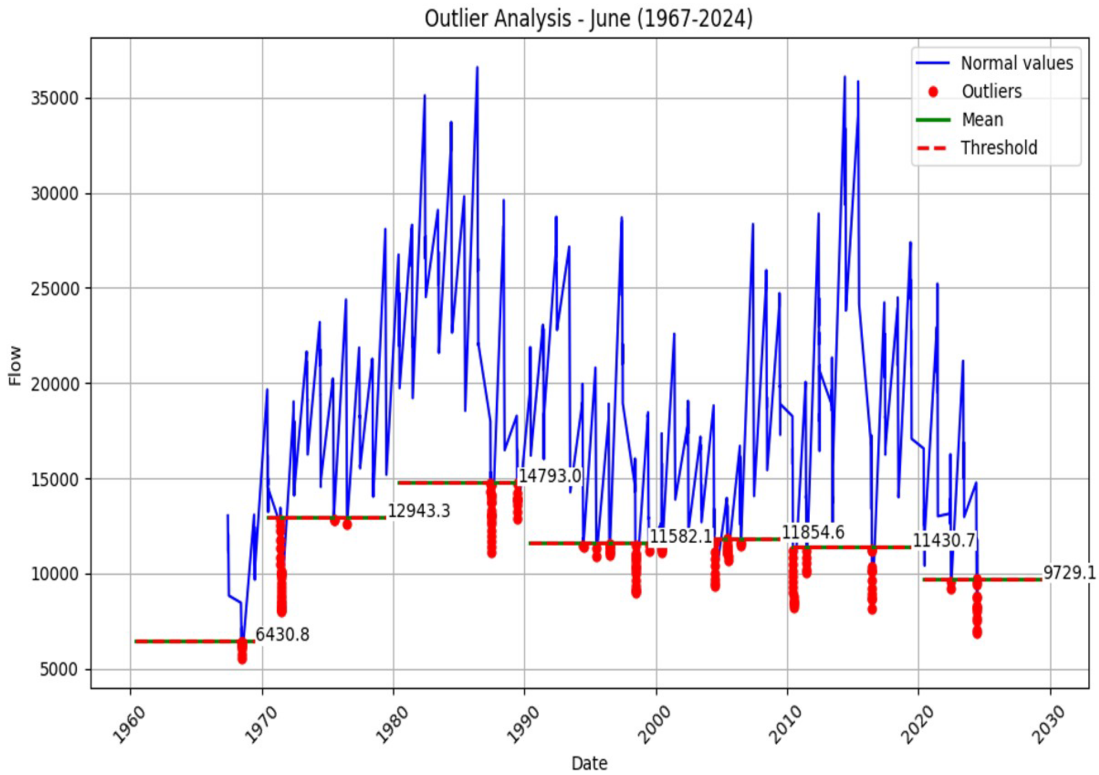


Figure 6. Dynamic flow thresholds (m³/s) for june, with the P10% threshold, scaled by month and decade

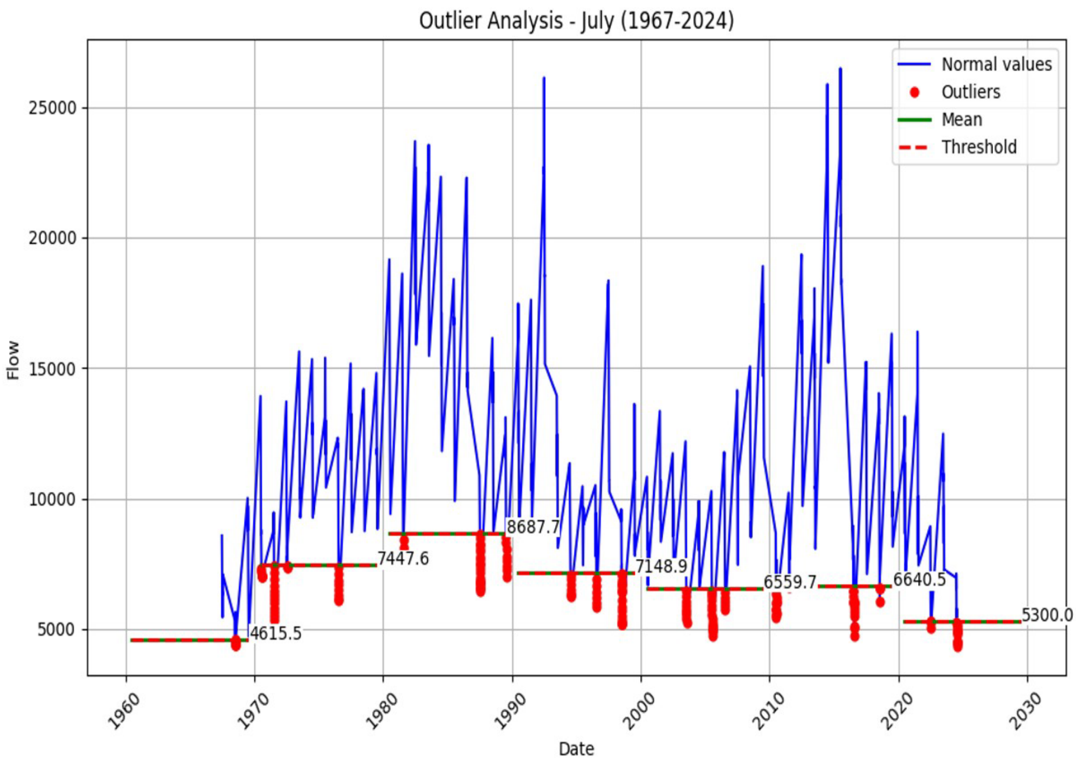


Figure 7. Dynamic flow thresholds (m³/s) for july, with the P10% threshold, scaled by month and decade

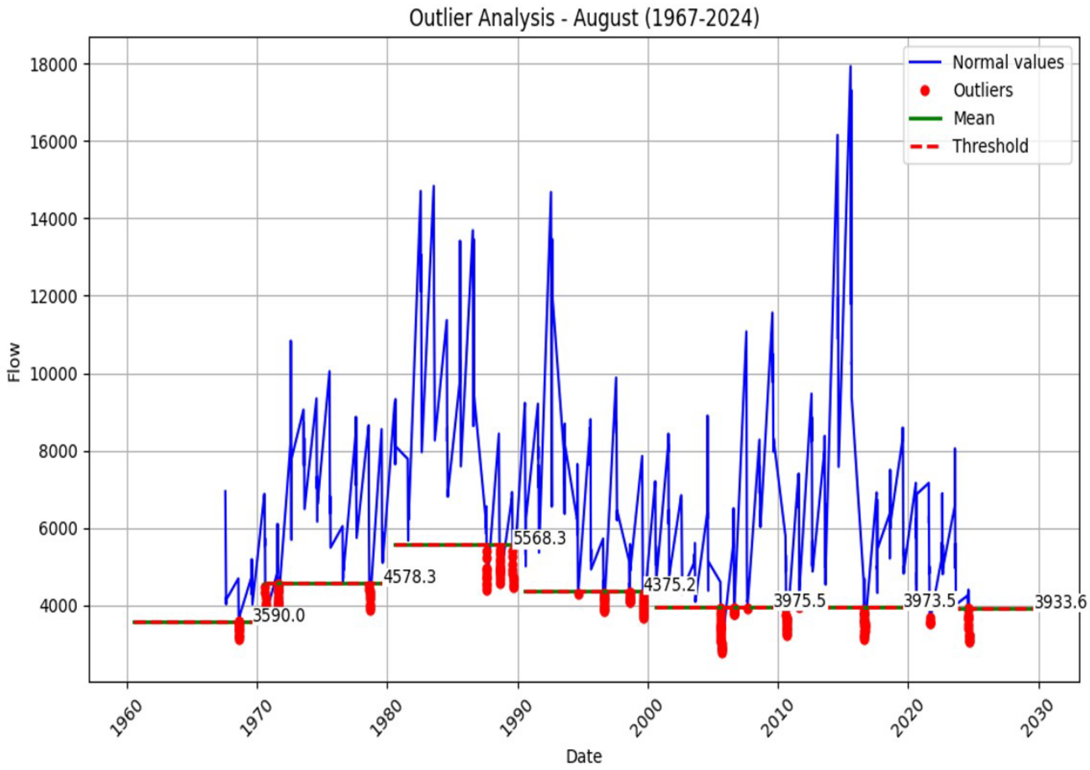


Figure 8. Dynamic flow thresholds (m³/s) for August, with the P10% threshold, scaled by month and decade

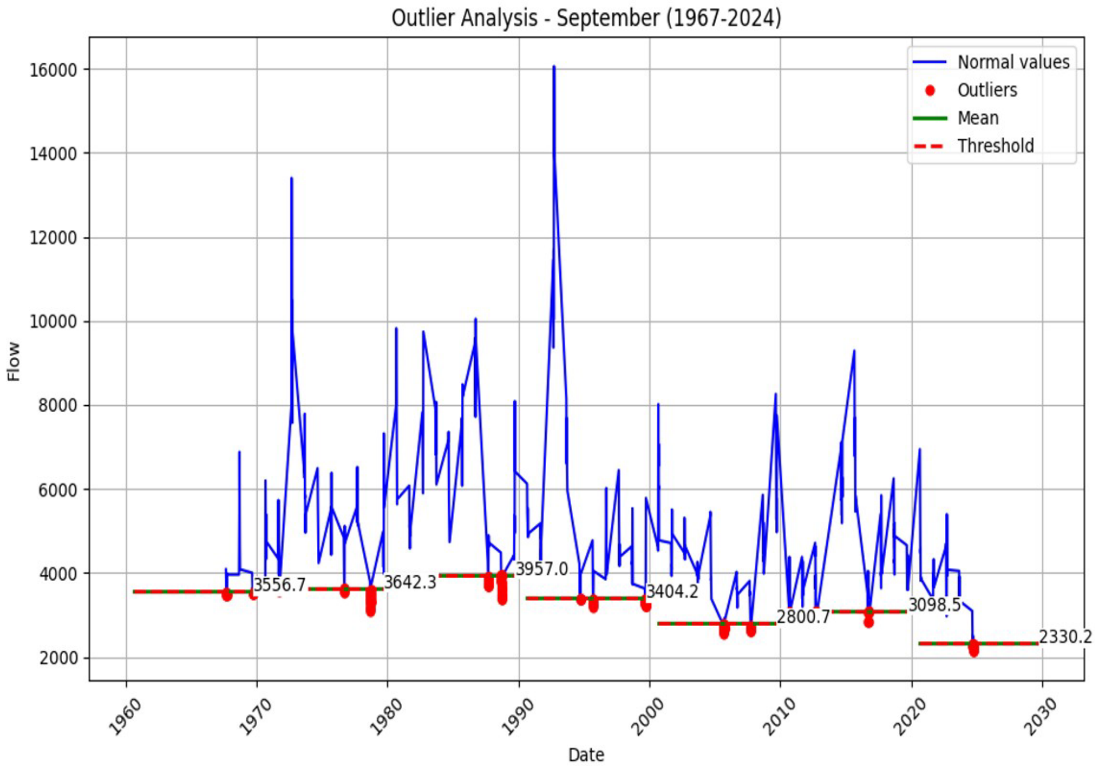


Figure 9. Dynamic flow thresholds (m³/s) for september, with the P10% threshold, scaled by month and decade

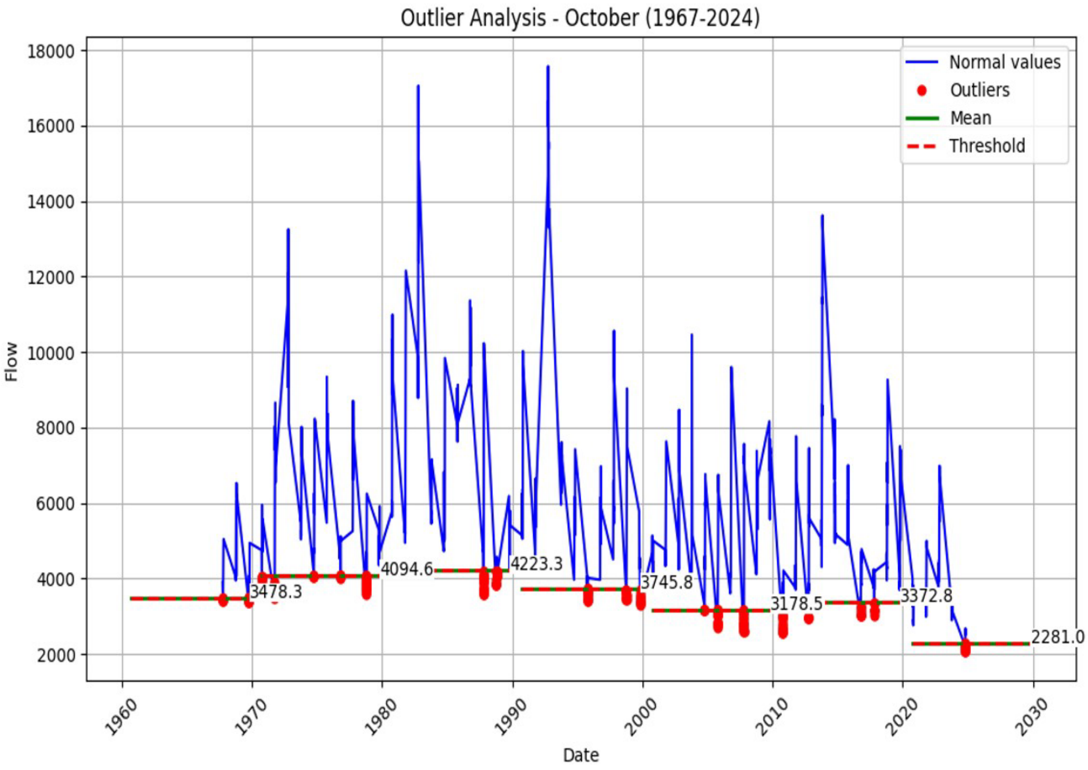


Figure 10. Dynamic flow thresholds (m³/s) for october, with the P10% threshold, scaled by month and decade

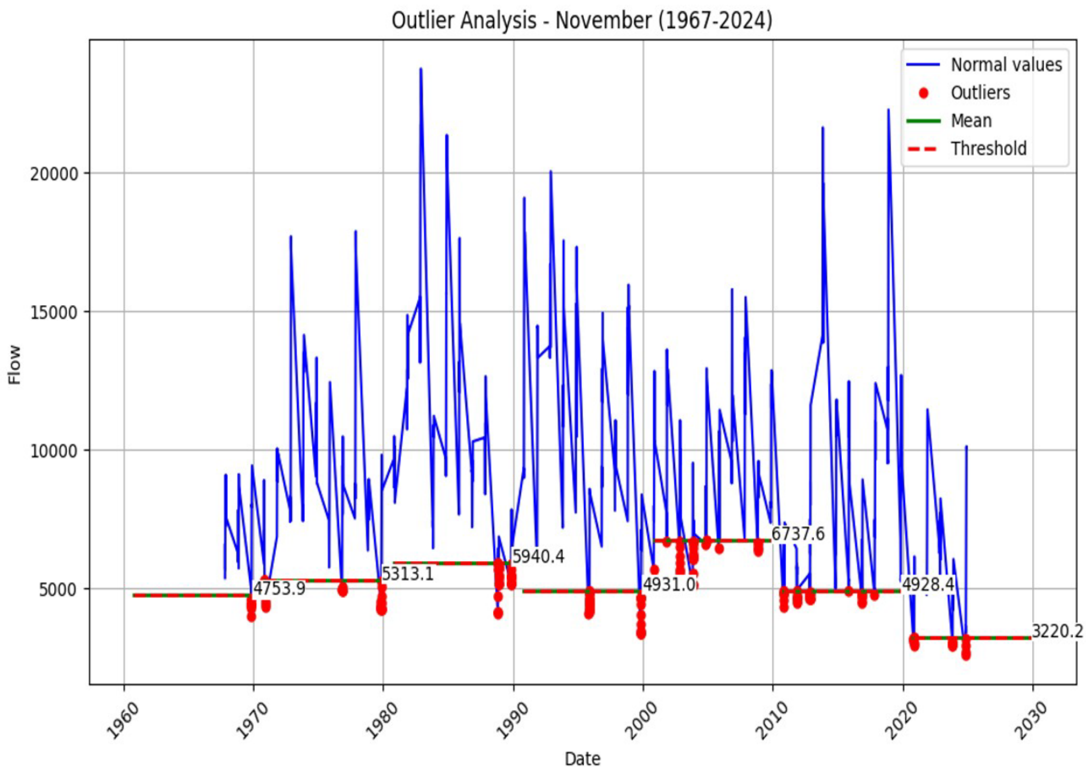


Figure 11. Dynamic flow thresholds (m³/s) for november, with the P10% threshold, scaled by month and decade

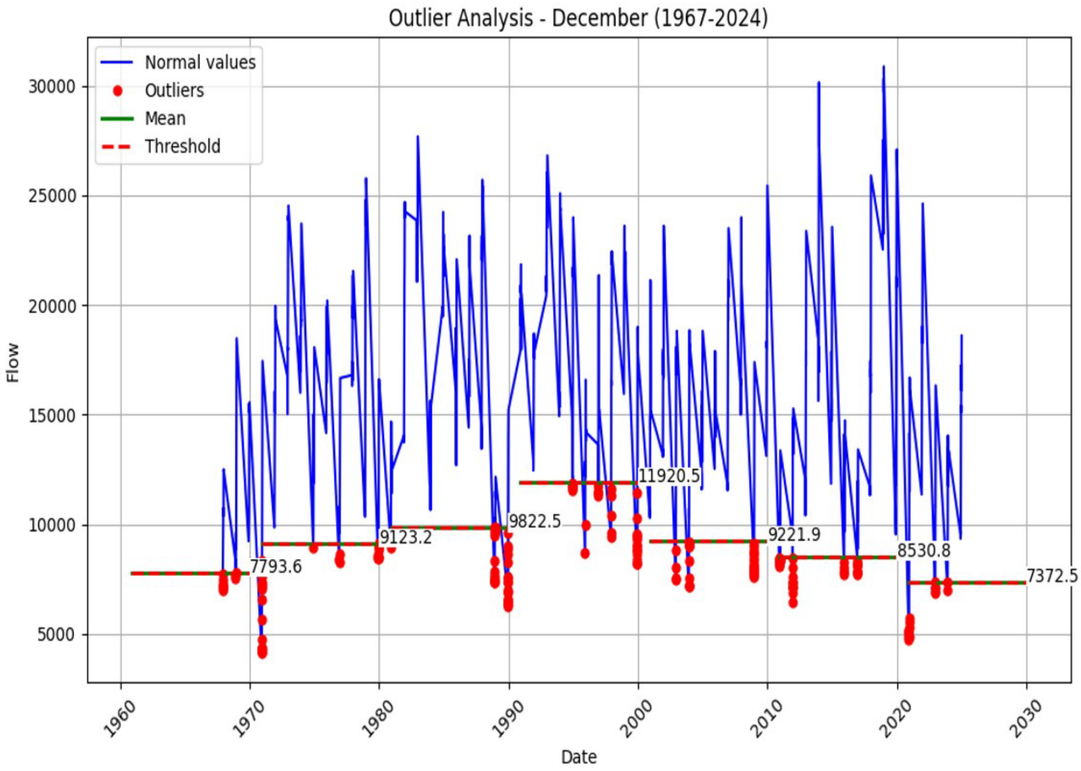


Figure 12. Dynamic flow thresholds (m³/s) for december, with the P10% threshold, scaled by month and decade

CONCLUSIONS

Monthly cross-validation has proven to be an effective tool for assessing the adequacy of the GEV distribution fit to monthly hydrological extremes. The results reveal significant seasonal variation both in the quality of fit and in the estimated GEV parameters, highlighting the need for flexible and period-specific approaches throughout the year. Therefore, the application of cross-validation techniques in extreme value studies is strongly recommended to ensure greater robustness and reliability in statistical inferences and hydrological risk management.

Conversely, the LOOCV method, by preserving each year as an independent unit for testing, provides a more rigorous and honest evaluation. The resulting higher KS statistics (≈0.50) and, more importantly, the substantially larger standard deviations, accurately reflect the model's sensitivity to the specific characteristics of individual years. This exposure of performance instability is not a flaw of the method but rather its principal strength; it yields a realistic, albeit more sober, estimate of the model's true predictive uncertainty when faced with unseen annual conditions.

Therefore, we conclude that for applications involving hydrological or other geophysical time-series data, validation methods that preserve the temporal data structure, such as LOOCV, are strongly preferred. The results from the LOOCV analysis are more defensible and provide a truer representation of the model's robustness and its limitations. The insights gained from this method are fundamental for establishing confidence bounds and understanding the real-world applicability of the statistical models.

REFERENCES

Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, *4*, 40-79. https://doi.org/10.1214/09-SS054

Beirlant, J., Goegebeur, Y., Segers, J., & Teugels, J. (2004). *Statistics of extremes: theory and applications.* Hoboken: Wiley. https://doi.org/10.1002/0470012382.

Coles, S. (2001). *An introduction to statistical modeling of extreme values.* London: Springer. https://doi.org/10.1007/978-1-4471-3675-0.

Embrechts, P.; Klüppelberg, C.; Mikosch, T. Modelling Extremal Events for Insurance and Finance. Berlin: Springer-Verlag, 1997.

Gumbel, E. J. *Statistics of Extremes*. New York: Columbia University Press, 1958.

Hosking, J. R. M., & Wallis, J. R. (1997). *Regional frequency analysis: an approach based on L-Moments.* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511529443.

Katz, R. W., Parlange, M. B., & Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources*, *25*(8-12), 1287-1304. https://doi.org/10.1016/S0309-1708(02)00056-8

Massey Junior, F. J. (1951). The kolmogorov-smirnov test for goodness of fit. *Journal of the American Statistical Association*, *46*(253), 68-78. https://doi.org/10.1080/01621459.1951.10500769

MathWorks. (2025). *Generalized extreme value distribution*. Retrieved in 2025, June 6, from https://www.mathworks.com/help/stats/generalized-extreme-value-distribution.html

NASA GMAO. (2025). *Generalized extreme value distribution*. Retrieved in 2025, June 6, from https://gmao.gsfc.nasa.gov/research/subseasonal/atlas/GEV-RV-html/GEV-RV-description.html

Renard, B., Lang, M., Bois, P., Dupeyrat, A., Mestre, O., Niel, H., Sauquet, E., Prudhomme, C., Parey, S., Paquet, E., Neppel, L., Gailhard, J. (2013). Regional methods for trend detection: assessing field significance and regional consistency. *Water Resources Research*, *49*(8), 4037-4050. https://doi.org/10.1029/2007WR006268.

Scarrott, C., & MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *Revstat Statistical Journal*, *10*(1), 33-60.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, *36*(2), 111-133. https://doi.org/10.1111/j.2517-6161.1974.tb00994.x

Author contributions:

Luiz Antônio de Oliveira: conducted the conceptualization, data processing, statistical analysis, model development, and manuscript writing. The author approves the final version of the manuscript.