

Intensity-Duration-Frequency Curves for Manicaragua city, Cuba

Roberto Luis López Ferras¹, Carlos Lázaro Castillo García², Ismabel Domínguez Hurtado³,
José Alejandro Solís⁴ y Lisdelys González Rodríguez⁵

Abstract — The Intensity-Duration-Frequency (IDF) curves are a way to visualize and represent extreme hydrometeorological rainfall events. In this article, an analysis of convective rainfall events recorded at the La Piedra Meteorological Station, Villa Clara, Cuba, was conducted. To develop IDF curves, the 2006-2019 time series was analyzed. A partial duration series was generated, including intervals from 20 minutes to 4320 minutes, subjected to an outlier detection process. The series was divided into two categories: one for durations ≤ 720 minutes and another for durations > 720 minutes. The resulting series underwent non-parametric tests to assess their independence, randomness, homogeneity, and seasonality. Subsequently, they were fitted to the Generalized Pareto probability distribution and to a parametric equation of the Montana model, and then the curves were plotted for return periods of 10, 50 and 100. The Montana model led to obtaining correlation coefficients greater than 0.90 compared to the other methods used, significantly improving the quality of the fit in both categories. This research provides information to understand and plan the management of intense climatic phenomena and adequate risk management in an area where such studies are lacking, facilitating access to crucial data essential in the design and execution of hydraulic engineering projects in the region.

Keywords – partial duration series, rainfall intensity, threshold, curves, precipitation.

Resumen — Las curvas de Intensidad-Duración-Frecuencia (IDF) son una manera de visualizar y representar los eventos hidrometeorológicos extremos de lluvia. En este artículo se llevó a cabo un análisis de eventos de lluvias convectivas registrados en la Estación Meteorológica La Piedra, Villa Clara, Cuba. Para desarrollar curvas IDF se analizó la serie temporal 2006-2019. Se generó una serie de duración parcial que incluyó intervalos de 20 minutos hasta 4320 minutos, sometiéndola a un proceso de detección de

datos anómalos. La serie se dividió en dos categorías: una para duraciones ≤ 720 minutos y otra para duraciones > 720 minutos. Las series resultantes se sometieron a pruebas no paramétricas para evaluar su independencia, aleatoriedad, homogeneidad y estacionalidad. Posteriormente, se procedió a ajustarlas a la distribución probabilística Generalizada de Pareto y a una ecuación paramétrica del modelo de Montana. Luego se grafica la representación de las curvas para períodos de retorno de 10, 50 y 100 años. El modelo de Montana condujo a la obtención de coeficientes de correlación superiores a 0.90 y a los demás métodos utilizados, mejorando significativamente la calidad del ajuste en ambas categorías. Esta investigación aporta información para comprender y planificar la gestión de fenómenos climáticos intensos, así como para una adecuada gestión de riesgos en una zona donde no se cuenta con este tipo de estudios. Asimismo, facilita el acceso a datos cruciales que resultan fundamentales en el diseño y ejecución de proyectos de ingeniería hidráulica en la zona.

Palabras claves – series de duración parcial, intensidad de la lluvia, umbral, curvas, precipitación.

I. INTRODUCTION

IN water resources management and engineering projects, Intensity-Duration-Frequency (IDF) curves are often needed. According to [1], the IDF curve establishes how the maximum annual average intensity varies in relation to the duration of a specific event. IDF curves are a widely used tool in hydrological design of maximum flows, when rainfall-runoff models such as unit hydrographs or the rational method are used [2]. Using this data, a large number of hydrological projects, such as flood discharge designs, bridge construction, and drainage network construction, are defined in relation to the maximum precipitation that could be expected for a given return period [2]. The parameters storm duration and intensity of the IDF curve are of great importance in the field of hydrology, as they are the basic elements when it comes to hydrological risk analysis [3]. According to [4], constructing the IDF relationship of rainfall is one of the main applications of Extreme Value Theory (EVT). IDF relationships are developed based on historical data from time series of rainfall by fitting a theoretical probability distribution to series of annual maximum values [5] (that is, maximum values per blocks with a block size of one year) or partial duration series (PDS), also known as peak over threshold series.

In the PDS approach, rainfall events with intensity exceeding the defined high threshold value are considered as the extreme rainfall series and are typically modeled using the Generalized Pareto Distribution (GPD) [6]. However, [7] in Israel analyzed IDF

1 A. Roberto Luis López Ferras is with the Central University Marta Abreu de Las Villas. Email: rlferraz@uclv.cu, ORCID: <https://orcid.org/0009-0001-4756-6496>

2 B. Carlos Lázaro Castillo García is with the Central University Marta Abreu de Las Villas. Email: cegarcia@uclv.cu, ORCID: <https://www.orcid.org/0000-0002-6430-2775>

3 C. Ismabel Domínguez Hurtado is with the Provincial Meteorological Center Villa Clara, Santa Clara, Cuba. Email: ismabel.dominguez@vcl.insmet.cu, ORCID: <https://www.orcid.org/0000-0002-7841-8031>

4 D. José Alejandro Solís is with the Central University Marta Abreu de Las Villas. Email: jsolis@uclv.cu, ORCID: <https://orcid.org/0009-0003-4777-605X>

5 E. Lisdelys González Rodríguez is with the Faculty of Engineering and Business, University of the Americas, Concepción Campus, Concepción 4030000, Chile. Email: lgonzalezr@udla.cl ORCID: <https://orcid.org/0000-0002-7892-4604>

Manuscript Received: 25/03/2024

Revised: 15/05/2024

Accepted: 30/05/2024

DOI: <https://doi.org/10.29019/enfoqueute.1046>

curves based on partial duration series and used both the Generalized Pareto Distribution (GPD) and the Generalized Extreme Value (GEV) distribution, in addition to Gumbel and Log normal distributions, finding that there are not many differences in the results. On the other hand, the extreme rainfall series obtained by extracting the maximum value in each calendar year is generally modeled using the Generalized Extreme Value (GEV) distribution, as seen in works by authors such as [8], [9], [10], and [11].

The annual maximum series (AMS) model shows a significant limitation as it does not consider secondary events within a year that may exceed the annual maxima of other years. The PDS overcomes the limitations of AMS by extracting all maximum values that exceed a particular discharge level, called a threshold [12]. When the threshold is appropriately selected, it can lead to reasonable variability in quantile estimates [13]. [14] demonstrated in their research that for records of less than 15 years, the PDS method is superior to the AMS method. On the other hand, [15] states that PDS is used when the data set is small (< 12 years) or when using return periods less than 5 years. [16] confirm this information when stating that for modeling return periods less than 10 years (2 and 5 years), PDS tends to produce higher values. This is supported by [17] in a comparison between PDS and AMS methodologies, demonstrating that for obtaining design storms, PDS is more effective than those imposed by the AMS Methodology since the rainfall exceeds by 4 to 10 %, and therefore is more conservative, and higher results are obtained for return periods between 2 and 5 years.

Despite the advantages shown by the PDS method, the difficulty in determining the appropriate threshold for PDS is a significant obstacle to its widespread use. As expressed by [18], there is no general recommendation for choosing a suitable threshold.

In their work, [19] present an equation that shows the relationship between the return period for AMS and PDS, as shown in the following expression (Eq. 1).

$$T_p = - \frac{1}{\ln \left(1 - \frac{1}{T_A} \right)} \quad (1)$$

Where TP is the return period of the PDS analysis and TA is the return period obtained from AMS. Even for return periods greater than 10 years, this expression can be reduced to $T_A = TP + 1/2$, which makes both results, for both PDS and AMS, relatively equal.

On the other hand, another aspect to consider is the variability in data series trends. [20] state that most efforts to develop future IDF curves are limited to individual cities or regions that assume stationarity in precipitation, works by [21] and [22] apply this methodology. Ganguli himself observes that for longer return periods (i.e., 25 years), detectable changes occur in the design storm considering a non-stationary model than the stationary ones. Studies conducted in Malaysia by [23] demonstrate that for non-stationary models with generalized extreme value functions, there are no clear advantages over similar stationary models.

The selection of the threshold in PDS is the main question that arises, there are several methods for this, but one has not yet been established that applies as a defined law for all cases.

[24] estimated a value of peaks per year (λ) ranging from 3 to 15 for homogeneous regions in Italy. Rosbjerg and Madsen (2004) cited in [25] conducted a frequency analysis based on rainfall data using a Bayesian framework in Denmark and suggested a value of λ between 2 and 3. Deidda (2010) cited by [26] developed a multiple threshold method, which is based on parameter estimates within a range of thresholds $u > u^*$ and provides a robust fit of the GPD regardless of data resolution or rounding. Regarding the choice of the optimal threshold u^* , it should be noted that it should be selected large enough to reliably consider that the distribution of exceedances approaches a GPD, but low enough to keep the estimation variance small. [27] conducted a comparison between PDS and AMS methods, where for obtaining the threshold they choose the minimum value of each duration from the AMS dataset to extract the PDS data. According to [28], since the selection of the threshold is an iterative process, one could easily choose any higher or even lower threshold value, but the lower threshold offers some advantages in terms of sample size. In their study, the author himself states that the results of statistical tests were considerably better if more data points are added. [29] predetermined the threshold values during the selection of rainfall data.

Human errors, such as data entry mistakes, can lead to the presence of outliers or atypical values in the data. Despite the ambiguity in providing a clear definition, an outlier is generally considered to be a data point that significantly differs from other data points or does not fit the expected normal pattern of the phenomenon it represents [30]. According to [31], they are data points that are extremely distant from the majority of other data points.

There are three types of outliers that a researcher may encounter [1]:

- Vertical outliers: These are extreme observations in the dependent variable.
- Good leverage points: These are extremes in both dependent and independent variables but fall on or near the regression line.
- Bad leverage points: These are outliers in the independent variables.

There is no solid theoretical basis to justify the choice of a specific probability distribution function, nor a theoretical procedure to define a probabilistic model as the best one in a frequency analysis contrasted with different probabilistic models [32]. The author himself established a definitive conclusion, stating that the competence of a probabilistic model in fitting hydrological data is directly related to the flexibility and form of the probability distribution function. Moreover, the more parameters a model contains, the more versatile its probability distribution function will be, adapting to the data.

The formulation proposed by Sherman in (1931), cited by [33], has universalized the mathematical and graphical representation for calculating intensity (i)-duration (d)-frequency (T) curves worldwide.

In Cuba, [34] conducted a study and proposed a methodology for creating IDF curves in our country. Subsequent research has been conducted, such as that by [35] and [36], both focusing on AMS. In this case, a different methodology is intended to be employed, as there is a small dataset of pluviographic records, specifically 13 years of records. The literature consulted

previously affirms that for return periods of less than 15 years, the PDS methodology is superior to AMS. With this, the aim is to establish a new methodology to apply in Cuba for meteorological stations with few years of records.

II. MATERIALS AND METHODS

A. Study Area

The “La Piedra” Meteorological Station (Code 78308) is located in Manicaragua, a municipality belonging to the province of Villa Clara, Cuba (Figure 1). This municipality is situated at latitude 22°9'0.8" N and longitude 79°58'43.2" W, nestled in the Escambray Mountains in the southern part of Villa Clara, bordering the provinces of Cienfuegos to the west and Sancti Spiritus to the east. It covers an area of 1066.0 km², making it the largest municipality in terms of surface area in the province, encompassing the southern portion, resembling a triangle with one of its vertices pointing south, at the intersection point of the provincial boundaries of Villa Clara, Cienfuegos, and Sancti Spiritus. Due to its geographical location, it experiences the influence of a warm humid tropical climate. Precipitation is regulated by the regime of the Northeast Trade Winds, which interact perpendicularly with the moist air masses from higher altitude regions, leading to intense processes with a marked increase in precipitation.



Fig. 1. a) Geographical location of the central region of Cuba, b) Area where the La Piedra Meteorological Station is located.

An analysis of the pluviographic data recorded during the period 2006-2019 (13 years) was conducted. A study of the maximum precipitation values at the La Piedra Meteorological Station was performed. The results are presented in the form of partial series, considering different time ranges: 20, 40, 60, 90, 120, 150, 240, 300, 720, 1440, 2880, and 4320 minutes.

It is important to mention that there are some limitations due to the lack of records for the year 2018, as the data recording sheets were lost. Additionally, records are occasionally interrupted for short periods due to breakdowns, maintenance, and instrument malfunctions. Another major limitation is the scarcity of recorded years, resulting from the few years the meteorological station has been in operation.

B. Flow Chart

The methodological scheme shown in Figure 2 can be applied to the development of IFD curves in stations that have analog and/or digital data records. It is divided into five stages and 14 phases.

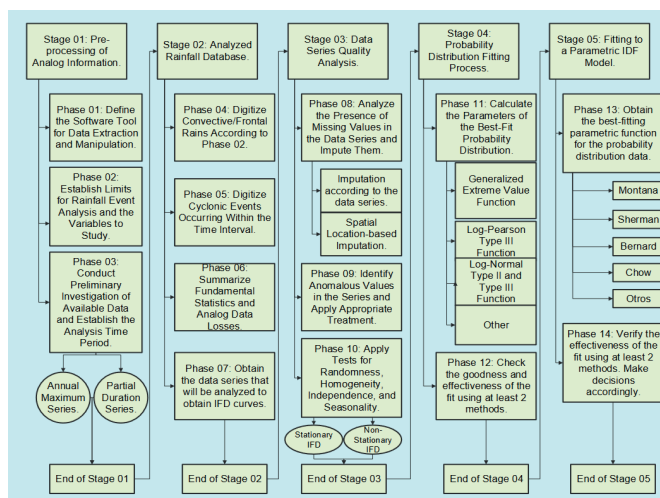


Fig. 2. Methodological scheme used in this study.

Stage 01. The partial duration series of precipitation were collected from the rainfall data recorded at the La Piedra Meteorological Station. Prior to this, the necessary conditions for the analysis of rainy events were established. To process the analog/digital information volumes, Excel, SPSS, and R software were employed.

Stage 02. The database of the selected rainfall parameters to be analyzed was designed with the aim of summarizing the essential components of the IFD analysis in each event, as well as other variables that may be of interest in other research. Since rainfall could be described in various ways as a completely random phenomenon, data were extracted from precipitation hyetographs or mass curves that reflect the statistical behavior of the rainfall shape, as reflected by [36]. It was necessary to select a threshold, for which a value was chosen in such a way as to satisfy the condition of independence of the series. A higher threshold value allows us to ensure greater probability of independence, but reduces the number of events in the series, which means a loss of valuable information. On the other hand, a lower threshold value provides a greater number of events in the series, thus allowing for a more reliable parameter estimation, but at the same time increases the possibility of independence. Therefore, a threshold was used that satisfies the condition of independence, while also allowing us to have the largest amount of data possible from the series. The procedure was carried out in such a way that a λ value between 2 and 3 was obtained. An attempt was made to obtain a number of peaks as close to 3 as possible in order to have a larger sample of data considering the limited number of years, as there are only thirteen years of records available.

Stage 03. Demonstrating the quality of the data obtained in the analysis of rainfall events was crucial to ensure the reliability of the predictions. In this regard, the treatment of anomalous data played a fundamental role in avoiding biases in the predictions. The application of sensitivity analysis methods in the analysis of outliers became an essential tool to reduce the uncertainty associated with an IFD curve study. For the analysis of anomalous hydrological data, the US-WRC (United States-Water Resources Council) method was employed. This method

is recommended by [37] and [38], where it is cautioned that to apply it, it is necessary to assume that the logarithms or another function of the hydrological series are normally distributed, as the test is only applicable to samples obtained from a normal population. Therefore, to implement the test, equations 2 and 3 are calculated:

$$X_H = \exp(\bar{x} + K_N s) \quad (2)$$

$$X_L = \exp(\bar{x} - K_N s) \quad (3)$$

Where, \bar{x} y s represent the mean and standard of the natural logarithms of the sample, respectively. The K_N statistic refers to the Grubbs and Beck table, which is tabulated for different sample sizes and levels of significance, and N denotes the sample size, while X_H is the upper limit of the test and X_L is the lower limit [37].

If $5 \leq N \leq 150$, K_N can be calculated using equation 4 (Stedinger and others, 1993 cited in [37]):

$$K_N = -0.9043 + 3.345\sqrt{\log(N)} - 0.4046 \log(N) \quad (4)$$

For the frequency analysis results to be valid, the dataset must meet the statistical criteria of randomness, independence, homogeneity, and stationarity [37]. Table 1 shows the test performed for each statistical criterion.

TABLE I
NON-PARAMETRIC TESTS FOR THE ANALYSIS
OF DATA QUALITY OF LA PIEDRA METEOROLOGICAL STATION

Statistical Criteria	Recommended Test	Confidence Interval (%)
Randomness	Run Test	95
Homogeneity	Mann-Withney test	
Independence	Wald-Wolfowitz test	
Seasonality	Mann- Kendall, Sen's Slope test	

The analysis of extreme hydrological data frequencies, such as floods, droughts, winds, and maximum daily precipitation, is based on accepting that the maximum annual data in the available sample are independent and come from a stationary random process [7]. The author himself expresses that due to some factors such as changes in land use and impacts of global warming, hydrological data series exhibit trends, which would indicate that they are not stationary. Therefore, it is necessary to check the existence of seasonality in the data series.

The Mann-Kendall test was developed to identify trends and analyze seasonality in data sets. This test, considered non-parametric, was designed to evaluate the presence of non-linear trends and change points in time series. It is commonly used in detecting trends and spatial variations in data related to climate, hydrology, and agrometeorology, as highlighted by [39].

The null hypothesis of trend, denoted as H_0 , holds that a sample of data arranged in chronological order is independent of each other and follows the same distribution at all time points. [40] defines the statistic S as presented in equations 5 and 6:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(x_j - x_i) \quad (5)$$

Where:

$$\text{sgn}(x) = \begin{cases} 1 & \text{si } x > 0 \\ 0 & \text{si } x = 0 \\ -1 & \text{si } x < 0 \end{cases} \quad (6)$$

When the sample size (n) is equal to or greater than 40, the statistic S follows an approximately normal distribution. In this case, the mean of that distribution is zero and the variance can be calculated using equation 7:

$$\text{Var}\{S\} = \frac{1}{18[n(n-1)(2n+5) - \sum_t t(t-1)(2t+5)]} \quad (7)$$

Where t is the size of a given tied group and \sum is the sum of the set of all tied groups in the data sample. The normalized test statistic K is calculated using the equation 8:

$$K = \frac{S-1}{\sqrt{\text{Var}(S)}}; \mathbf{0}; \frac{S+1}{\sqrt{\text{Var}(S)}} \text{ para: } \mathbf{S} > \mathbf{0}, \mathbf{S} = \mathbf{0}, \mathbf{S} < \mathbf{0}, \quad (8)$$

The statistic K follows a standard normal distribution, which means it has a specific distribution shape with a mean of 0 and a variance of 1. To find the probability (P) associated with the statistic K in your sample data, you can use the cumulative distribution function of the normal distribution, in the form of equation 9:

$$P = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{t^2}{2}} dt \quad (9)$$

When analyzing sets of independent data without showing any trend, the value of P should hover around 0.5. In situations where the data exhibit a clear trend towards increasing values, the value of P tends to approach 1. On the other hand, if the trend is marked in a downward direction, the value of P approaches 0. If the sample data are serially correlated, it will be necessary to whiten the data beforehand and apply a correction to calculate the variance [37].

When we are seeking linear trend in a dataset, we commonly use the least squares estimation technique via linear regression to calculate the slope. However, this approach is reliable only if there is no systematic correlation between consecutive observations (serial correlation). Additionally, it is important to note that this method can be heavily influenced by outliers in the data, meaning that an unusual data point can have a significant impact on the slope estimation. Sen (1968) developed a more robust method [36], [37].

The slope of the trend was calculated as equation 10:

$$\beta = \text{mediana} \left(\frac{x_i - x_j}{i - j} \right); \forall j < i \quad (10)$$

The trend slope β was estimated using the formula where Q is the resulting value and x_i and x_j represent the data values at

times i and j . If β is positive, it indicates an upward trend, while if it is negative, it signals a downward trend.

The Sen's slope estimator is simply the median of the N' values of β . This approach is applied in the same way, whether we have one or multiple observations per time period.

Sen (1968) proposed a non-parametric method to calculate a confidence interval for the slope. However, in practice, a simpler method based on the normal approximation is more commonly used. To calculate this, we require the standard deviation of the Mann-Kendall statistic, S [37]. In simpler terms, we need to know how much the Mann-Kendall statistic S varies in order to perform calculations related to its distribution and statistical significance.

Stage 04. Identifying the probability function best fitting the data for the evaluated conditions and the proposed significance level was essential. Fit was assessed using analytical and probabilistic methods to ensure the correct choice of the distribution function and the results of its parameters.

In the field of hydrology, probability distributions are essential tools used to analyze precipitation amounts and patterns in time series. To describe rainfall events, three distributions were employed: (1) Generalized Extreme Value (GEV), (2) Generalized Pareto [29], (3) Johnson SB. Equations 11, 12, and 13 show the probability density function of GEV, the Generalized Pareto distribution, and the probability density function (Johnson, 1949), [34], respectively:

$$F(x; \mu, \sigma, \xi) = \exp \left[- \left(1 + \xi^* \left(\frac{x-\mu}{\sigma} \right)^{\frac{1}{\xi}} \right) \right] \quad (11)$$

Where μ is the location parameter, σ is the scale parameter, and ξ is the shape parameter. When $\xi=0$, the GEV reduces to the type I extreme value distribution (Gumbel). When $\xi>0$, the tail of the distribution is heavier, and when $\xi<0$, the tail is lighter.

$$FX(x) = 1 - [1 - \kappa(x-\xi)/\alpha]^{1/\kappa} \quad (12)$$

Where κ is a shape parameter, ξ is the location parameter, and α is the scale parameter.

$$f(x) = \frac{\delta}{\sqrt{2\pi}} \frac{\lambda}{(x-\varepsilon)(\varepsilon+\lambda-x)} e^{\frac{1}{2}[\gamma+\delta \ln(\frac{x-\varepsilon}{\varepsilon+\gamma-x})]^2} \quad (13)$$

ε, λ : Location and scale parameters.

γ, δ : Shape parameters representing skewness and kurtosis, respectively.

Parameter estimation A very tempting method from a statistical point of view is the maximum likelihood method. It consists of selecting the parameters that give a fitting distribution the greatest statistical coherence possible with the observed sample.

The method of maximum likelihood is the most commonly used method to find the parameter values that make the observed data most probable under the proposed model, as affirmed by Millar (2011) and Pawitan (2001), both cited by [24]. The author himself expresses that the method provides biased estimates, hence researchers strive to develop nearly unbiased estimators for the parameters of various distributions.

Goodness of fit [30] stated that many of the problems observed in numerous research studies stem from the incorrect use of test statistics, which occurs when they are applied inappropriately. This can be due to various factors:

- Lack of knowledge of both descriptive and inferential statistics.
- Limited understanding of research methodology.
- Shortage of research-oriented faculty members.
- Inadequate familiarity with statistical software.

These four aspects directly influence the conduct of research. If any of them is not understood or applied properly, the research would exhibit significant deficiencies. [37] explains that in hydrology, there are various rigorous and effective statistical tests to evaluate whether it is plausible to conclude that a specific set of observations comes from a particular distribution. These tests are called Goodness of Fit Tests. The Kolmogorov-Smirnov test allows for obtaining bounds for each of the observations on a probability plot when the sample has been effectively drawn from the assumed distribution.

This procedure is a non-parametric test that allows testing whether two samples come from the same probabilistic model. Suppose we have two samples of total size $N=m+n$ composed of observations $x_1, x_2, x_3, \dots, x_n$ e $y_1, y_2, y_3, \dots, y_m$. The test assumes that the variables x e y are mutually independent and that each x comes from the same continuous population P_1 and each y comes from another continuous population P_2 . The null hypothesis is that both distributions are identical, meaning they are two samples from the same population [36].

The Kolmogorov-Smirnov test was used due to its ability to reasonably assume that observations could follow the specific distribution in question. It is straightforward to calculate and apply, and it does not require data grouping, unlike the Chi-square test. Additionally, it has the advantage of being applicable to samples of any size, unlike the Chi-square test, which requires a minimum sample size.

Stage 05. The results of the probability function were parameterized into mathematical models of the form $f(I)=[D]$ for different T , or $f(I)=[D; T]$, in order to verify the effectiveness of the fits. If necessary, a point of inflection in the data series was identified, contributing to obtaining more precise results with less margin of error. Models for Intensity-Frequency-Duration curves: In this study, we focused on fitting the models proposed by Montana, Sherman, Bernard, and Chow according to [33], [36], equations 14-17.

$$i_d^T = \frac{kT^m}{d^{\theta+C}} \quad (14)$$

$$i_d^T = \frac{kT^m}{(d+C)^n} \quad (15)$$

Where: i_d^T It represents the maximum precipitation intensity, measured in mm/min or mm/h. T is the return period in years, d is the duration of precipitation in minutes, and k, m, θ and n are the parameters that must be estimated to fit the curve.

III. RESULTS AND DISCUSSION

The Table II shows the selection of thresholds is shown below, indicating the threshold and the number of resulting peaks in each of the time series.

TABLE II
RESULTS OF THRESHOLD SELECTION
AND NUMBER OF PEAKS FOR EACH SERIES

Series min	Threshold mm/min	N ^o of peaks per year
20	1	2.92
40	0.75	2.77
60	0.55	2.69
90	0.43	2.85
120	0.33	3.0
150	0.28	2.92
240	0.185	3.0
300	0.15	3.0
720	0.069	3.0
1440	0.038	2.92
2880	0.025	3.0
4320	0.0185	3.0

The data processing was done with the assistance of the R software. The US-WRC method was applied, which extracts the outlier data from each series. Table III displays these outlier data.

TABLE III
ANOMALOUS DATA EXTRACTED FROM EACH OF THE SERIES

Series (min)	Anomalous Data (mm/min)	Date
20	3.5	23/07/2019
40	2.46	23/07/2019
60	1.5	07/04/2019
	1.83	23/07/2019
90	1.02	30/06/2006
	2.01	24/05/2012
	1.27	07/04/2019
	1.31	23/07/2019
	0.76	30/06/2006
120	0.67	09/09/2008
	1.5	24/05/2012
	1.17	07/04/2019
	0.99	23/07/2019
150	0.61	30/06/2006
	0.58	09/09/2008
	1.2	24/05/2012
	1.1	07/04/2019
	0.79	23/07/2019
240	0.51	09/09/2008
	0.75	24/05/2012
	0.72	07/04/2019
	0.49	23/07/2019

300	0.49	09/09/2008
	0.6	24/05/2012
	0.57	07/04/2019
720	0.39	23/07/2019
	0.386	09/09/2008
	0.258	24/05/2012
	0.423	09/09/2017
1440	0.238	07/04/2019
	0.193	09/09/2008
	0.129	24/05/2012
2880	0.212	09/09/2017
	0.158	28/05/2018
	0.145	09/09/2008
	0.149	09/09/2017
4320	0.143	28/05/2018
	0.1068	09/09/2008
	0.0678	24/05/2012
	0.1176	09/09/2017
	0.1255	28/05/2018

The data shown in Table III were removed from the series for exceeding 10 % above the upper limits of the US-WRC model for a 95 % confidence level.

The results of the quality tests conducted on the partial duration series as recommended by the [36] included Runs, Mann-Whitney (M-W), Wald-Wolfowitz (W-W), and Mann-Kendall (M-K) tests and are summarized in Table IV. The expression ‘OK’ means that the null hypothesis is accepted, and the word ‘NO’ means that the null hypothesis is not accepted, indicating that:

1. The series is random at a significance level of 5 % (Runs Test).
2. The series is independent at a significance level of 5 % (Wald-Wolfowitz Test).
3. The series is homogeneous at a significance level of 5 % (Mann-Whitney Test).
4. The series is seasonal at a significance level of 5 % (Mann-Kendall Test).

TABLE IV
RESULTS OF QUALITY TESTS FOR EACH DATA SERIES

Series (min)	Rachas	M-W	W-W	M-K
20	OK	OK	OK	NO
40	OK	OK	OK	NO
60	OK	OK	OK	NO
90	OK	OK	OK	NO
120	OK	OK	OK	NO
150	OK	OK	OK	NO
240	OK	OK	OK	NO
300	OK	OK	OK	NO
720	OK	OK	OK	NO
1440	OK	OK	OK	NO
2880	OK	OK	OK	NO
4320	OK	OK	OK	NO

According to the results obtained in the quality tests, it has been demonstrated that the data series collected for the La Piedra Meteorological Station are suitable for use in the development of Intensity-Duration-Frequency (IDF) curves. These results highlight the data's capability to be used in probabilistic analyses, suggesting the suitability of applying non-stationary models in the processing and interpretation of such information, in order to understand and predict climate patterns with greater accuracy and relevance.

Figure 3 shows the results of the analysis using the Generalized Pareto distribution, which has been adjusted using the nearest neighbor method. The results are presented for intervals of 1h, 2h, 4h, and 12h. The adjustment corresponding to the previously mentioned durations is illustrated using the cumulative probability function graph.

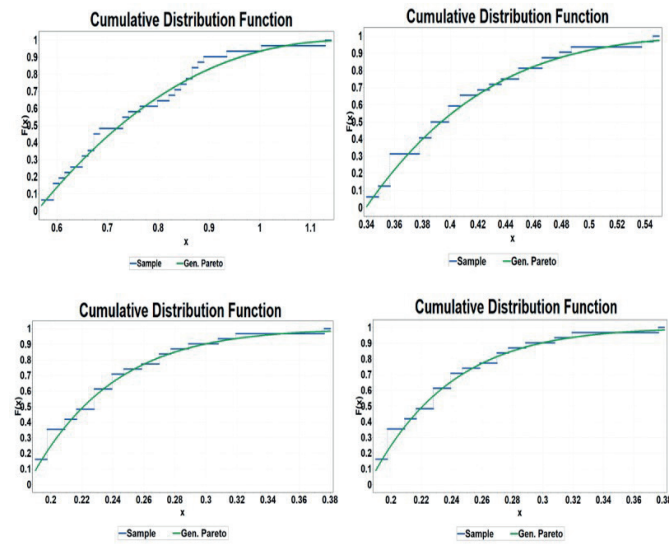


Fig. 3. Fit to the Generalized Pareto cumulative probability function for intervals of a) 1h, b) 2h, c) 4h, d) 12h.

In Table V, the results of the position and scale parameters derived from this analysis are detailed.

TABLE V

PARAMETERS OF THE GENERALIZED PARETO PROBABILITY DISTRIBUTION OBTAINED FOR THE 1H, 2H, 4H, 12H SERIES

Series	Location parameter μ	Scale parameter σ	Shape parameter k
1 hora	0.56029	0.27309	-0.42739
2 horas	0.33943	0.08459	-0.2314
4 horas	0.18496	0.05312	-0.06513
12 horas	0.06739	0.02166	-0.0024

The Kolmogorov-Smirnov goodness-of-fit test is conducted to assess the suitability of the (1) Generalized Pareto, (2) Generalized Extreme Value ($\xi=0$), (3) Johnson SB functions in obtaining the optimal probability function. With a significance level of 95 %, the test results indicate that the fit by the Generalized Pareto is statistically more effective.

The application of the Montana, Sherman, Bernard, and Chow models to fit the results of the Pareto probability function in a partial duration series did not yield favorable results. In this context, we rely on the Pearson correlation coefficient, detailed in Table VI, along with the values of k , m , θ , C , and n associated with each model.

TABLE VI

PARAMETERS OBTAINED AND PEARSON CORRELATION RESULTS FOR THE APPLIED MODELS

As a result, the suggestion is to divide the series into two segments: the first covering up to 12 hours and the second beyond 12 hours. By implementing this modification in the models, Table VII presents new values for k , m , θ , C , and n , along with a new Pearson correlation coefficient.

TABLE VII

PARAMETERS OBTAINED AND PEARSON CORRELATION RESULTS FOR THE APPLIED MODELS WITH DURATIONS LESS THAN AND GREATER THAN 12 HOURS

Model	k	m	θ	c	n	Pearson
Montana (-12 h)	93.946	0.111	1.09	46.734	-	0.9250
Montana (+12h)	$3.1 \cdot 10^{10}$	0.261	3.28	$7.4 \cdot 10^{11}$	-	0.9863
Sherman (-12h)	108.149	0.111	-	34.12	1.111	0.9236
Sherman (+12h)	2.351	0.258	-	$-5 \cdot 10^{-5}$	0.553	0.9519
Bernard (-12 h)	9.749	0.111	-	-	0.664	0.9108
Bernard (+12 h)	2.351	0.258	-	-	0.553	0.9519
Chow (-12h)	24.125	0.132	-	0.182	-	0.6480
Chow (+12h)	35.532	0.367	-	0.018	-	0.9354

According to the data presented in Table VII, it is concluded that the Montana model proves to be the most suitable for fitting the results of the Generalized Pareto probability distribution. This analysis supports the choice of the Montana model as the one that best fits the data of the series in question, and Equation 18 represents the conclusive formula used for the analysis at the corresponding station.

$$I = \left\{ \begin{array}{l} \frac{93.946T^{0.111}}{(D^{1.093} + 46.734)} \text{ for } D \leq 720 \text{ min} \\ \frac{30993903192.321T^{0.261}}{(D^{3.280} + 737208928517.207)} \text{ por } D > 720 \text{ min} \end{array} \right\} \quad (18)$$

Model	k	m	θ	c	n	Pearson
Montana	77.823	0.111	1.055	36.949	-	0.6581
Sherman	75.725	0.111	-	29.262	1.045	0.6969
Bernard	9.894	0.111	-	-	0.669	0.6875
Chow	26.101	0.124	-	0.123	-	0.6652

In Figure 4, the graphical representation of the Montana fitting function specifically for the La Piedra Meteorological Station is visually presented.

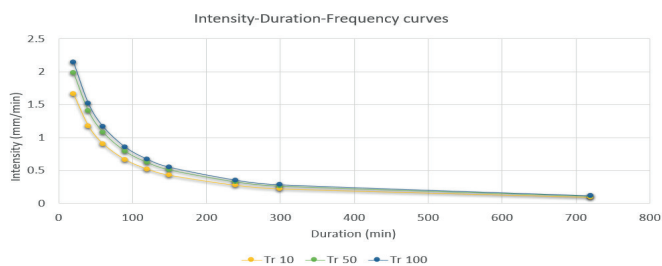


Fig. 4. Graph of Intensity-Duration-Frequency (IDF) curves for the La Piedra Meteorological Station.

Table VIII shows the numerical percentage difference between the values obtained using the Montana method and the values from NC 1239-2018. For this purpose, the results of the Montana method were compared with those of NC 1239-2018, calculating the difference between both sets of data.

TABLE VIII
COMPARISON BETWEEN THE RESULTS OBTAINED
BY THE MONTANA METHOD AND NC 1239-2018

Series (min)	Return period		
	10 years	30 years	100 years
20	41.2 %	98 %	139 %
40	27.4 %	69 %	102 %
60	27 %	63 %	92 %
90	28 %	59 %	86 %
120	29 %	57 %	81 %
150	30 %	55.3 %	78.2 %
240	29 %	50.2 %	70 %
300	27 %	47 %	65.9 %
720	22 %	36 %	49.6 %

The difference was expressed as a percentage to clearly identify the maximum deviation between the compared values. It was determined that the greatest discrepancy, 139 %, was observed for a return period of 100 years and a rainfall duration of 20 minutes.

IV. CONCLUSIONS

After conducting the study and analysis of the data obtained throughout this research, the following conclusions are drawn:

- The collected data cover a period of 13 years of pluviographic records from the La Piedra Meteorological Station. During the processing of this data, series representing partial duration for intervals of 20, 40, 60, 90, 120, 150, 240, 300, 720, 1440, 2880, and 4320 minutes were generated.
- Pluviographic records for the year 2018 were not available due to total loss.
- 41 values considered outliers were identified using the US-WRC method. These values were excluded from our

analysis due to the lack of information about their origin or source.

- The data series underwent analysis to verify their randomness, independence, homogeneity, and seasonality. As a result of this analysis, the hypotheses were corroborated, and it was concluded that a non-stationary IDF model would accurately represent the studied phenomenon.
- The Pareto distribution was applied to the data series, and it was verified, through the Kolmogorov-Smirnov method, that this distribution fits the data adequately. This fit was highlighted when compared to the Johnson SB and Generalized Extreme Value distributions, with the Pareto distribution yielding the most favorable results in comparison.
- The Montana model achieved parameterization with a more significant correlation when applied to data from the probability distribution. It is important to note that this was achieved by dividing the series into two categories: durations less than 720 minutes and durations greater than this figure. This division led to the formulation of a Montana model with two equations presenting different parameters.
- The results of this research present data developed through the Montana method, which, when compared with NC 1239-2018, showed a difference in values exceeding 100 %.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to the Provincial Meteorological Center of Villa Clara for their invaluable collaboration in the development of this research. The support provided has been exceptional and deserves special recognition.

REFERENCES

- [1] R. Balbastre Soldevila, *Análisis comparativo de metodologías de cálculo de tormentas de diseño para su aplicación en hidrología urbana*. [Tesis de Maestría, Universitat Politècnica de València], 2018. Available: <http://hdl.handle.net/10251/100090>
- [2] P. L. Martínez Rodas, *Curvas de Intensidad-Duración-Frecuencia para la ciudad de Cuenca*. [Magíster en Higiene y Sanidad, Universidad del Azuay], 2023. Available: <http://dspace.uazuay.edu.ec/handle/datos/12941>
- [3] A. G. Yilmaz, H. Safaet, F. Huang and B. J. C. Perera, "Time-varying character of storm intensity frequency and duration curves," *Australasian Journal of Water Resources*, vol. 18, no. 1, 15-26, 2014, <https://doi.org/10.7158/W12-017.2014.18.1>
- [4] V. Agilan, and N. V. Umamahesh, "What are the best covariates for developing non-stationary rainfall Intensity-Duration-Frequency relationship?," *Advances in Water Resources*, vol. 101, 11-22, 2017c. <https://doi.org/https://doi.org/10.1016/j.advwatres.2016.12.016>
- [5] M. Noor, T. Ismail, E.-S. Chung, S. Shahid and J. H. Sung, "Uncertainty in Rainfall Intensity Duration Frequency Curves of Peninsular Malaysia under Changing Climate Scenarios," *Water*, vol. 10, no. 12, 2018, <https://doi.org/https://doi.org/10.3390/w10121750>
- [6] V. Agilan and N. V. Umamahesh, "Non-Stationary Rainfall Intensity-Duration-Frequency Relationship: a Comparison between Annual Maximum and Partial Duration Series," *Water Resources Management*, vol. 3, no. 6, 1825-1841, 2017b, <https://doi.org/10.1007/s11269-017-1614-9>

- [7] D. F. Campos-Aranda, "Ajuste con momentos L de las distribuciones GVE, LOG y PAG no estacionarias en su parámetro de ubicación, aplicado a datos hidrológicos extremos," *Agrociencia*, vol. 52, no. 2, 169-189, 2018. Available: https://www.scielo.org.mx/scielo.php?pid=S1405-31952018000200169&script=sci_arttext
- [8] S. Emmanouil, A. Langousis, E. I. Nikolopoulos and E. N. Anagnostou, 2020. Quantitative assessment of annual maxima, peaks-over-threshold and multifractal parametric approaches in estimating intensity-duration-frequency curves from short rainfall records. *Journal of Hydrology*, vol. 589, 125151, <https://doi.org/https://doi.org/10.1016/j.jhydrol.2020.125151>
- [9] P. Coelho Filho, J. Alexandre, D. C. de Rezende Melo and M. de Lourdes Martins Araújo, Estudio de chuvas intensas para a cidade de Goiânia/GO por meio da modelação de eventos máximos anuais pela aplicação das distribuições de Gumbel e Generalizada de Valores Extremos. *Ambiência*, vol. 13, no. 1, 2017. Available: <https://core.ac.uk/download/pdf/230459134.pdf>
- [10] C. Montesinos, W. Lavado, N. Quijada, L. Gutierrez and O. Felipe, Desarrollo de curvas pluviométricas Intensidad-Duración-Frecuencia (IDF) en Perú. Servicio Nacional de Meteorología e Hidrología del Perú- SENAMHI, 2023. Available: <https://repositorio.senamhi.gob.pe/handle/20.500.12542/2825>
- [11] J. L. Ng, S. K. Tiang, Y. F. Huang, N. I. F. M. Noh and R. A. Al-Mansob, "Analysis of annual maximum and partial duration rainfall series," *IOP Conference Series: Earth and Environmental Science*, vol. 646, no. 1, 012039, 2021, <https://doi.org/10.1088/1755-1315/646/1/012039>
- [12] S. Swetapadma and C. S. P. Ojha, "Chapter 9 - A comparison between partial duration series and annual maximum series modeling for flood frequency analysis," in K. S. Kasiviswanathan, B. Soundharajan, S. Patidar, J. He and C. S. P. Ojha (eds.), *Developments in Environmental Science*, vol. 14, 173-192, Elsevier, 2023, <https://doi.org/https://doi.org/10.1016/B978-0-443-18640-0.00007-9>
- [13] C. Leys, M. Delacre, Y. L. Mora, D. Lakens and C. Ley, "How to classify, detect, and manage univariate and multivariate outliers, with emphasis on pre-registration," *International Review of Social Psychology*, vol. 32, no.1, 2019, <https://doi.org/http://doi.org/10.5334/irsp.289>
- [14] W. Martín Rosales, A. Pulido Bosch, Á. Vallejos and M. López Chicano, "Precipitaciones máximas en el Campo de Dalias y vertiente meridional de la Sierra de Gador (Almería)," *Geogaceta*, vol. 20, no. 6, 1251-1254, 1996. Available: <https://dialnet.unirioja.es/servlet/articulo?codigo=8115318>
- [15] G. Zucarelli, N. Piccoli, M. Pittau and M. Gallo, "Curvas intensidad-duración-frecuencia en la Región Litoral de la República Argentina," *Cuadernos del CURIHAM*, vol. 15, no. 0, 69-76, 2009, <https://doi.org/10.35305/curiham.v15i0.71>
- [16] A. G. Yilmaz, H. Safaet, F. Huang and B. J. C. Perera, "Time-varying character of storm intensity frequency and duration curves," *Australasian Journal of Water Resources*, vol. 18, no. 1, 15-26, 2014, <https://doi.org/10.7158/W12-017.2014.18.1>
- [17] S. Vrbán, Y. Wang, A. McBean Edward, A. Binns and B. Gharabaghi, Evaluation of stormwater infrastructure design storms developed using partial duration and annual maximum series models," *Journal of Hydrologic Engineering*, vol. 23, no. 12, 04018051, 2018, [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001712](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001712)
- [18] N. Guru and R. Jha, "A Framework for the Selection of Threshold in Partial Duration Series Modeling," In R. Jha, V. P. Singh, V. Singh, L. B. Roy and R. Thendiyath (eds.), *Hydrological Modeling: Hydraulics, Water Resources and Coastal Engineering* (pp. 69-84), 2022. Springer International Publishing. https://doi.org/10.1007/978-3-030-81358-1_7
- [19] W. Martín Rosales, A. Pulido Bosch, Á. Vallejos and M. López Chicano, "Precipitaciones máximas en el Campo de Dalias y vertiente meridional de la Sierra de Gador (Almería)," *Geogaceta*, vol. 20, no. 6, 1251-1254, 1996. Available: <https://dialnet.unirioja.es/servlet/articulo?codigo=8115318>
- [20] P. Ganguli and P. Coulibaly, Assessment of future changes in intensity-duration-frequency curves for Southern Ontario using North American (NA)-CORDEX models with nonstationary methods. *Journal of Hydrology: Regional Studies*, vol. 22, 100587, 2019, <https://doi.org/https://doi.org/10.1016/j.ejrh.2018.12.007>
- [21] M. T Vu, V. S. Raghavan and S. Y. Liong, "Deriving short-duration rainfall IDF curves from a regional climate model," *Natural Hazards*, vol. 85, no. 3, 1877-1891, 2017, <https://doi.org/10.1007/s11069-016-2670-9>
- [22] J. Li, J. Evans, F. Johnson and A. Sharma, "A comparison of methods for estimating climate change impact on design rainfall using a high-resolution RCM," *Journal of Hydrology*, vol. 547, 413-427, 2017, <https://doi.org/https://doi.org/10.1016/j.jhydrol.2017.02.019>
- [23] M. Noor, T. Ismail, E.-S. Chung, S. Shahid and J. H. Sung, "Uncertainty in Rainfall Intensity Duration Frequency Curves of Peninsular Malaysia under Changing Climate Scenarios," *Water*, vol. 10, no. 12, 2018, <https://doi.org/https://doi.org/10.3390/w10121750>
- [24] P. Claps and F. Laio, "Can continuous streamflow data support flood frequency analysis? An alternative to the partial duration series approach," *Water Resources Research*, vol. 39, no. 8, 2003, <https://doi.org/https://doi.org/10.1029/2002WR001868>
- [25] N. Guru and R. Jha, "A Framework for the Selection of Threshold in Partial Duration Series Modeling," In R. Jha, V. P. Singh, V. Singh, L. B. Roy and R. Thendiyath (eds.), *Hydrological Modeling: Hydraulics, Water Resources and Coastal Engineering* (pp. 69-84), 2022. Springer International Publishing. https://doi.org/10.1007/978-3-030-81358-1_7
- [26] S. Emmanouil, A. Langousis, E. I. Nikolopoulos and E. N. Anagnostou, 2020. Quantitative assessment of annual maxima, peaks-over-threshold and multifractal parametric approaches in estimating intensity-duration-frequency curves from short rainfall records. *Journal of Hydrology*, vol. 589, 125151, <https://doi.org/https://doi.org/10.1016/j.jhydrol.2020.125151>
- [27] V. Agilan and N. V. Umamahesh, "Non-Stationary Rainfall Intensity-Duration-Frequency Relationship: a Comparison between Annual Maximum and Partial Duration Series," *Water Resources Management*, vol. 3, no. 6, 1825-1841, 2017b, <https://doi.org/10.1007/s11269-017-1614-9>
- [28] F. Karim, M., Hasan and S. Marvanek, "Evaluating Annual Maximum and Partial Duration Series for Estimating Frequency of Small Magnitude Floods," *Water*, vol. 9, no. 7, 4812017, <https://doi.org/10.3390/w9070481>
- [29] J. L. Ng, S. K. Tiang, Y. F. Huang, N. I. F. M. Noh and R. A. Al-Mansob, "Analysis of annual maximum and partial duration rainfall series," *IOP Conference Series: Earth and Environmental Science*, vol. 646, no. 1, 012039, 2021, <https://doi.org/10.1088/1755-1315/646/1/012039>
- [30] H. Wang, M. J. Bah and M. Hammad, 2019, Progress in Outlier Detection Techniques: A Survey. *IEEE Access*, 7, 107964-108000, <https://doi.org/10.1109/ACCESS.2019.2932769>
- [31] C. Leys, M. Delacre, Y. L. Mora, D. Lakens and C. Ley, "How to classify, detect, and manage univariate and multivariate outliers, with emphasis on pre-registration," *International Review of Social Psychology*, vol. 32, no.1, 2019, <https://doi.org/http://doi.org/10.5334/irsp.289>
- [32] W. T. Hernández Guarín and P. X. Moreno Vivas, Regionalización de sequía hidrológica en la cuenca del río Bogotá a partir del método de l-momentos, 2017. Available: <http://hdl.handle.net/11634/9266>
- [33] A. Gutiérrez-López and R. Barragán-Regalado, "Ajuste de curvas IDF a partir de tormentas de corta duración," *Tecnología y ciencias del agua*, vol. 10, no. 6, 1-24, 2019. <https://doi.org/10.24850/j-tyca-2019-06-01>
- [34] Y. Rodríguez López, N. Marrero de León and A. León Méndez, "Consideraciones prácticas sobre las curvas IFD," *Ingeniería Hidráulica y Ambiental*, vol. 30, no. 1, 2009. Available: <https://link.gale.com/apps/doc/A304466968/IFME?u=anon-91a3f3a4&sid=googleScholar&xid=c12dbfc5>
- [35] S. Barcia Sardiñas and O. González, "Determinación de la curva de intensidad-duración-frecuencia de Cienfuegos," *Revista Cubana De Meteorología*, 19(1), 3-12, 2013. Available: <http://rcm.insmet.cu/index.php/rcm/article/view/140>
- [36] C. Castillo-García, I. Domínguez-Hurtado, Y. Martínez-González and D. Abreu-Franco, "Curvas de intensidad-duración-frecuencia para la ciudad de Santa Clara, Cuba," *Tecnología y Ciencias del Agua*, vol. 15, no.1, 361-408, 2024, <https://doi.org/10.24850/j-tyca-15-01-09>
- [37] OMM, Guía de prácticas hidrológicas, Gestión de Recursos hídricos y aplicación de prácticas hidrológicas, 2011, Sexta edición ed., Vol. II. Available: https://library.wmo.int/doc_num.php?explnum_id=10038

- [38] M. Naghettini, *Fundamentals of statistical hydrology*. Springer, 2017. <https://doi.org/https://doi.org/10.1007/978-3-319-43561-9>
- [39] S. F. A. Xavier Júnior, J. d. S. Jale, T. Stosic, C. A. C. d Santos and V. P. Singh, "Precipitation trends analysis by Mann-Kendall test: a case study of Paraíba, Brazil," *Revista Brasileira de Meteorología*, vol. 35, 2020, <https://doi.org/https://doi.org/10.1590/0102-7786351013>
- [40] R. Maity. *Statistical methods in hydrology and hydroclimatology*. Springer Singapore, 2018, <https://doi.org/https://doi.org/10.1007/978-981-16-5517-3>