

Article

Maximum Extreme Flow Estimations in Historical Hydrological Series under the Influence of Decadal Variations

Marco Antonio Jacomazzi ¹, Antonio Carlos Zuffo ¹, Monzur Alam Imteaz ^{2,*} , Vassiliki Terezinha Galvão Boulomytis ^{1,3}, Marcus Vinícius Galbetti ¹ and Tais Arriero Shinma ⁴

¹ School of Civil Engineering, Architecture and Urban Design, State University of Campinas, Campinas 13083-852, Sao Paulo, Brazil; marcoantonio@rasa.eng.br (M.A.J.); zuffoac@unicamp.br (A.C.Z.); vassiliki@ifsp.edu.br (V.T.G.B.); mvgalbetti@uems.br (M.V.G.)

² Department of Civil & Construction Engineering, Swinburne University of Technology, Hawthorn, VIC 3122, Australia

³ Faculty of Civil Engineering, Federal Institute of Science, Education and Technology of Sao Paulo, Caraguatuba 11665-071, Sao Paulo, Brazil

⁴ School of Environmental Engineering, State University of Mato Grosso do Sul, Dourados 79804-970, Brazil; taisarriero@uems.br

* Correspondence: mimteaz@swin.edu.au

Abstract: The hypothesis of stationarity is a fundamental condition for the application of the statistical theory of extreme values, especially for climate variables. Decadal-scale fluctuations commonly affect maximum and minimum river discharges. Thus, the probability estimates of extreme events need to be considered to enable the selection of most appropriate time series. The current study proposed a methodology to detect the fluctuation of long wet and dry periods. The study was carried out at the gauging station 4C-001 in Pardo River, State of São Paulo, Brazil. The Spearman, Mann–Kendall and Pettitt’s non-parametric tests were also performed to verify the existence of a temporal trend in the maximum annual daily flows. The graph achieved from the Pettitt’s statistical variable allowed for the identification and separation of the longest dry period (1941 to 1975) and the longest wet period (1976 to 2011), decreasing again in 2012. Analysing the series separately, it was observed that both mean and standard deviation were higher than those corresponding to the dry period. The probable maximum flows for the corrected series showed estimates 10% higher than those estimated for the uncorrected historical series. The proposed methodology provided more realistic estimates for the extreme maximum flows.

Keywords: extreme events; decadal fluctuation; non-parametric tests; hydrological time series; probability distributions



Citation: Jacomazzi, M.A.; Zuffo, A.C.; Imteaz, M.A.; Boulomytis, V.T.G.; Galbetti, M.V.; Shinma, T.A. Maximum Extreme Flow Estimations in Historical Hydrological Series under the Influence of Decadal Variations. *Hydrology* **2022**, *9*, 130. <https://doi.org/10.3390/hydrology9080130>

Academic Editor: Yanfang Sang

Received: 27 June 2022

Accepted: 22 July 2022

Published: 25 July 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Extreme events are the most impacting phenomena on society and infrastructure worldwide, and they present the greatest challenges for managers and technicians to design and implement mitigation measures. Information on the impact and trends of these events can improve the estimates of future events and can thus be used to support political decisionmaking and flood management and to control system design [1].

The importance of hydro-climate estimates is demonstrated in the large number of scientific papers that have been published recently aimed at the improvement of probabilistic estimates [2–5]. Comparative studies of parameter estimation methods have been widely used to analyse hydro-climate extremes. The need to improve these estimates is supported by the application of new mathematical tools to describe natural phenomena such as artificial intelligence [6]. Table 1 presents three more studies about the subject. In the study of Ribeiro Junior et al. [7], normal distribution is applied to the precipitation series by the use of nonparametric methods, which determine the non-homogeneity in rainfall series and subsidize rainfall–runoff modelling of large basins.

The homogeneity of natural climate time series has also gained prominence regarding identifying decadal variations in rainfall series [1,8–10] or heat flows from the Antarctic seas associated with the El Niño Southern Oscillation [11], which influences global climate.

The fluctuation in precipitation regimes is a complicated phenomenon that can cause significant impacts on environmental and socioeconomic components (e.g., disrupting water supply and demand, impacting vegetation, and causing floods or droughts) [8–12]. Identifying the pattern of decadal precipitation fluctuations is critical because it can help us manage water supply and mitigate natural disasters such as floods and droughts, which often occur around the world [13]. As precipitation estimates are notoriously uncertain, particularly for extreme local events, a probabilistic method is generally used to simulate possible scenarios. Despite the calculation method, rainfall frequency analyses are used for the estimation of design rainfalls [14], which are totally related to the infrastructure design in urban areas. Flash floods are urban hazards and are associated with extreme short-term rainfalls in daily or hourly scales [15–17]. They occur due to the low capacity of the catchments to accumulate runoff discharge [18].

Table 1. Comparative studies assessing different parameter estimation methods for the analyses of hydro-climate extremes.

Reference	Variable	Location (Number of Sites)	Sample Size (Years)	Distributions or Other Methodologies	Methods	Evaluation Criteria
[19]	floods	Pakistan (18)	not available	GLO, GEV, GPA	LH-moments, Goodness-of-Fit	Anderson Darling, Kolmogorov–Smirnov, Cramér–Von Mises, <i>t</i> -test
[20]	Precipitation	Crete (Greece) (54)	41	GAM, GEV, LGN, WBL, P-II	KDE, BGK	KCDE-ITS, CDF, ERA5—Copernicus Climate Change Service
[7]	Precipitation and average river flow	Brazil (22)	51	ND	MGB-IPH, ETA model/A1B/IPCC	Mann–Kendall, Pettitt, Run, Spearman

One of the major issues of our day is predicting extreme hydrological occurrences in the long and medium terms under non-stationary settings [21]. Runoff discharge and temporal rainfall are among the hydrological phenomena that can be considered as stochastic processes with random variables within a given probability distribution, for which time series demonstrations are widely used [21,22]. As per Koutsoyiannis and Montanari [23], the concept of stationarity is used to describe a probability density function which may not be changed in time. Thus, the calculation of distribution quantiles (i.e., division of a probability distribution into equal probability areas) associated with a specific non-exceedance probability is one of the challenges of extreme event frequency analysis. They are frequently calculated when a probabilistic model is fitted to observable data [21].

According to Totaro et al. [21], nonparametric measures of trends have been used for the detection of non-stationary conditions in extreme events, including statistical tests such as Mann–Kendall [24] and Spearman [25] as well as the Pettitt’s [26] tests for change point detection. They are all founded on a specific null hypothesis to achieve a certain significance level. However, despite its nonparametric structure, Mann–Kendall test effectiveness depends on the kind and parameterization of the parent distribution [27].

Due to being distribution-free, not requiring extensive knowledge about the parent distribution and being less sensitive to the presence of outliers, these non-parametric tests are frequently favoured over parametric tests for the frequency analysis of extreme events [28,29].

2. Materials and Methods

In the study of the variation of extreme flows, we used the maximum annual flows from the hydrological history of daily flows. For the study, we selected the gauging station 4C-001 in Pardo River, which is part of the Mogi-Guaçu River catchment, in the City of Ribeirao Preto, State of São Paulo, Brazil.

This hydrological series was selected because it has good quantitative and qualitative historical data (i.e., updated extent of historical series, no incomplete years of observations, and only a few gaps for filling or with non-consistent information). The location, analysed period, and watershed area are shown in Table 2.

Table 2. Gauging station features.

Gauging Station (Code/Name/City)	Area ¹ (km ²)	Analysed Period (Years/Series)	Coordinates
4C-001 Clube de Regatas Ribeirao Preto	10.679	1941–2021 81	S 21°6'12" W 47°45'27"

¹ Watershed area upstream of the gauging station.

Daily flow rate updates are the responsibility of the Hydraulic Technology Centre (CTH) of the University of Sao Paulo (USP), operated by the Department of Water and Electric Energy of the State of Sao Paulo (DAEE).

Selecting the “ideal size” of a hydrological series is challenging, especially when the data are scarce, but the main criterion considered for this study was to choose a long and good-quality data series. If it is not long enough, it becomes more complex to acknowledge the hydrological behaviour of the local extreme events. The fluctuation evaluations show the distinguished climate features of extreme events. This procedure is essential in estimating design rainfall and water availability, among other input parameters used by hydrologists to produce outcomes and to support efficient infrastructure design and urban planning.

Initially, the study of the hydrological series comprised qualitative evaluation using descriptive statistics (averages and standard deviations) and by the analysis of the periods of high and low peaks of the hydrograph. Then, we evaluated if the hydrological historical series was representative, random and without statistically significant trends. To assess whether the variations in the hydrological historical series occurred randomly, the non-parametric Wald–Wolfowitz statistical test was applied, known as the run test (oscillations). If the run test is being tested for randomness, then it is understood that the data should come into the dataset as an ordered sample, increasing in magnitude.

According to Ribeiro Junior [30], the run test counts the number of oscillations of the values above the reference measure (median) (i.e., positive, and negative values, respectively), in a naturally ordered data series. This number of oscillations is called “Run”, and the accounting of each oscillation is computed when there is a change between the positive and negative oscillations, or viceversa. For instance, when a value above the median (positive oscillation) is preceded by another, also oscillating in the positive range, it is not considered a variation of the oscillation and therefore is not counted as a run. However, if the preceding value is below the median (negative oscillation) it shows the occurrence of a variation and therefore is considered a run. If the number of values, either above the median (positive variations, $N1$) or below it (negative variations, $N2$) is greater than 20, the sample is considered large, and the run value sample distribution approaches the normal distribution. The expected value (E) and variance (Var) are calculated by Equations (1) and (2), respectively.

$$E(ru) = 1 + \frac{2 \times N1 \times N2}{N1 + N2} \quad (1)$$

$$Var(ru) = \frac{2 \times N1 \times N2 \times (2 \times N1 \times N2 - N1 - N2)}{(N1 + N2)^2 \times (N1 + N2 - 1)} \quad (2)$$

in which ru is the number of oscillations or runs as well as the statistical value of the test, $N1$ is the number of observations above the median (positive variations), and $N2$ is the number of observations below the median (negative variations).

Once the test distribution is approaching the normalization, the reduced variable value of the test is calculated by Equation (3). Adopting a significance level of 5%, the null hypothesis ($H0$ for a random sample series) will not be rejected if the calculated value (Z , reduced variable) is between $[-1.96; +1.96]$.

$$Z = \frac{u - E(u)}{\text{Var}(u)^{0.5}} \quad (3)$$

in which Z is the value of the reduced variable to the standard normal distribution and is derived by the statistical test, and u is the statistical value of the hypothesis test.

To assess whether there was a temporal trend in the historical hydrological series of the maximum annual daily flows (Qd_{max}), we combined the non-parametric statistical tests of Spearman (SP), Mann–Kendall (MK) and Pettitt (PT).

According to Pellegrino [31], to estimate the sequential test of MK, the tn statistic is initially calculated by Equation (4), considering the time series of the Yi variable (maximum annual daily flows, Qd_{max}) of N observations (years). This equation consists of determining the sum of the number of terms (mi) in the series, relative to the value Yi , whose preceding terms ($j < i$) account for the number of values $Yj < Yi$.

$$tn = \sum_{i=1}^n mi \quad (4)$$

in which tn is the value of the MK test, and mi is the number of preceding observations that satisfy the condition $Yj < Yi$, where $j < i$.

We used Equation (3) for the significance test of tn , represented as $Z(tn)$, which is the two-sided test for normal distribution. The expected value of the test and the variance were calculated by Equations (5) and (6), respectively.

$$E(tn) = \frac{N \times (N - 1)}{4} \quad (5)$$

$$\text{Var}(tn) = \frac{N(N - 1)(2N + 5)}{72} \quad (6)$$

in which $E(tn)$ is the expected value of the MK test with $\text{Var}(tn)$ being the variance of the MK test.

The calculated value of $Z(tn)$ is compared to the reduced variable of the normal distribution (z) extracted from the statistical tables and associated with the significance levels of: (i) 10% (significant), where $z = 1.64$, and (ii) 5% (highly significant), where $z = 1.96$.

The test hypotheses are: (i) $H0$, if there is no time trend, and (ii) Ha if there is a time trend. The purpose of the test is to verify whether the null hypothesis can be accepted or rejected, and thus, the module of the Z_{MK} value must not exceed the tabulated one.

To confirm the non-existence of a monotonous temporal trend for the hydrological series, the Spearman's non-parametric test was simultaneously employed. This test estimates the correlation coefficient between the series and the time index [32]. The value of the test statistic (rs) is calculated by Equation (7), in which the order of the increasing position of the values of the series (mi) is subtracted by the temporal position (time index, Ti).

$$rs = 1 - \frac{6 \times \sum_{i=1}^N (mi - Ti)^2}{N^3 - N} \quad (7)$$

where mi is the position of the observed event, regarding the increasing order of the values of the hydrological sample series, Ti is the sequential time composition of the event, and N is the number of observations of the sample series (number of observed years).

For large samples, r_s approaches the reduced normal distribution and is calculated by Equation (3). The expected value is null, and the test variance is calculated by Equation (8). For a significance level of 5%, if the calculated Z-value is between $[-1.96; +1.96]$, the null hypothesis is not rejected, and the series is considered stationary. The variance of the SP test is derived from Equation (8).

$$Var(rs) = \frac{1}{(N - 1)} \tag{8}$$

The point of abrupt change or beginning of the trend was determined by the non-parametric Pettitt’s test. This test verifies that two samples of the same series belong to the same population: (i) X_i, \dots, X_{t-1} and (ii) X_t, \dots, X_n .

The statistics $U(t,N)$ of this test are calculated by Equation (9), which is the total number of times that the observation of the first sample is larger than the members of the second. The variable $U(t,N)$ is calculated sequentially for all values of the hydrological series, as presented by Equation (9).

$$U(i, T) = U(i - 1, N) + \sum_{j=1}^T sgn(y_i - y_j) \tag{9}$$

where,

$$\begin{cases} sgn(y_i - y_j) \geq 0 \rightarrow +1 \\ sgn(y_i - y_j) = 0 \rightarrow 0 \\ sgn(y_i - y_j) \leq 0 \rightarrow -1 \\ i > j \end{cases}$$

in which $U(i,N)$ is the Pettitt’s statistic value for the observation y_i , $U(i-1,N)$ is the Pettitt’s statistic value for the previous observation y_{i-1} , and y_i is the value of the observation in the i order of the historical series. The variable $U(t,N)$ is calculated sequentially for all values of the hydrological series, in which $k(t)$ is the test value being the absolute maximum, as presented by Equation (10).

$$k(t) = Maximo_{i < t < T} |U(t, T)| \tag{10}$$

According to Back [33], the test significance is estimated by Equation (11):

$$p \cong 2 \times exp \left[\frac{-6 \times k(t)^2}{(N^3 + N^2)} \right] \tag{11}$$

in which p is the test significance for the desired significance level (α_o).

Based on Equation (11) and isolating the value of $k(t)$, we obtain Equation (12), associating the probability of significance level of $\alpha_o = 5\%$, and we calculate the critical value of the Pettitt’s statistic, which is k_{crit} .

$$k_{crit} \cong \mp \sqrt{\frac{-Ln\left(\frac{\alpha_o}{2}\right) \times (N^3 + N^2)}{6}} \rightarrow \mp \sqrt{\frac{-Ln\left(\frac{0,05}{2}\right) \times (N^3 + N^2)}{6}} \tag{12}$$

To prove that there was some abrupt change in the hydrological series and that it is not stationary, the module of the $k(t)$ statistic should be greater than the k_{crit} [33].

Once the stationary nature of the river flow series has been confirmed, we proceed to identify the periods with distinct hydrological behaviour in relation to the estimates of the sample mean and standard deviation. These estimates were prepared for two periods: (i) the beginning of the hydrological series from 1941 to 1975, which corresponds to the long dry period, and (ii) between 1975 and 2020, which is the long wet period. The first interval represents the series featured by the driest period, which comprises the lowest

observed maximum flows. In the following period, there was a change in both sample mean and standard deviation.

Once a periodic variability was identified in the hydrological series of maximum annual daily flows, these values were corrected, eliminating the temporal variation. For that, we used the methodology adapted from Detzel et al. [34] in the current study.

On the accumulative curve of the hydrological variable, time is represented on the horizontal axis, and the sum of the maximum annual flow is on the vertical axis. According to Detzel et al. [34], this method is based on the fact that by plotting the graph of the accumulative curve of the hydrological variable, it is possible to adjust the linear regression of continuous trend over the time horizon. In the study, a non-stationary pattern was identified by the change in curve slope, where it was possible to divide it into two linear regressions.

The correction for non-stationarity was carried out by adjusting the angular coefficients of the two fitted straight lines:

- The historical series was divided into two periods, and the limit of these new series was defined by the year of the maximum value of the statistic indicated by the PT test.
- The angular coefficients were estimated for both periods of the divided series.
- To linearize the trend, the Qd_{max} of the oldest period (i.e., the beginning of the series related to the rupture year) was multiplied by the relation of the linear coefficients of the recent period over the previous one.

The result of the correction is a random, stationary series in which the previous period (beginning of the series related to the year of rupture) is increased by the angular coefficient of the recent period, in which the observed values remain. The corrected series eliminates the long dry period, maximizing the estimates of design flows for hydraulic works.

To estimate the probability of occurrence of extreme maximum events, we used the Fisher–Tippet Type I distribution, commonly referred to as Gumbel distribution. The form of this probability function (FD) is shown in Equation (13).

$$FD(y < Y) = 1 - \exp\left(-e^{-\left(\frac{Yi-\alpha}{\beta}\right)}\right) \quad (13)$$

where $FD(Yi)$ is the cumulative likelihood function for the random variable.

The parameters α and β were estimated by the maximum likelihood method, as suggested by Beijo [35] and Bautista [36] and as shown in Equations (14) and (15), respectively.

$$\hat{\beta} = Ym - \frac{\sum yi \times e^{-yi/\hat{\beta}}}{\sum e^{-yi/\hat{\beta}}} \quad (14)$$

$$\hat{\alpha} = \hat{\beta} \times \text{Ln} \left[\frac{\sum e^{-yi/\hat{\beta}}}{N} \right] \quad (15)$$

in which Ym is the sample mean of the random variable, β is the scale parameter, and α is the position parameter.

Finally, we used the Kolmogorov–Smirnov test as described in Naghettini and Pinto [32] to assess whether the data fitted the theoretical probability distribution.

Figure 1 presents the flowchart of the methodology developed in this study.

For the estimation of the non-parametric statistical tests and adjustment of the Gumbel probability distribution parameters, scripts of these subroutines were developed in VBA (Visual Basic Application) programming language, implemented in the calculation platform of the Microsoft EXCEL®.

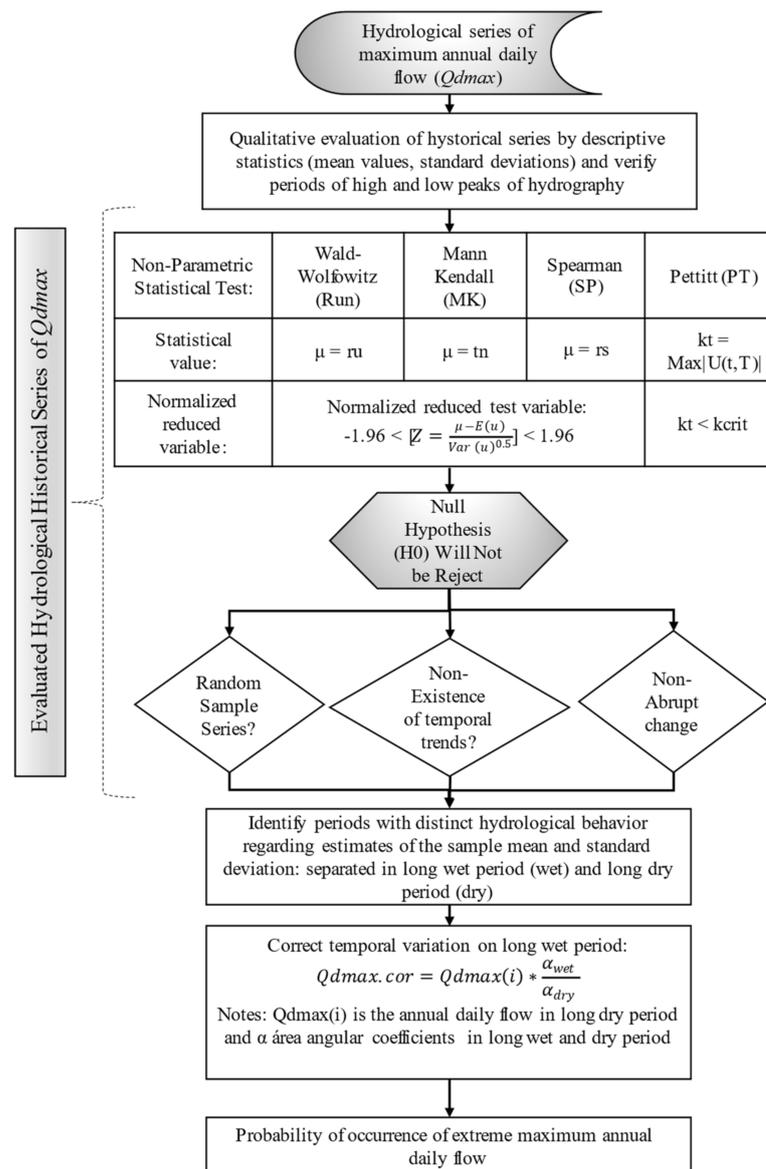


Figure 1. Flowchart of the methodology developed in this study.

3. Results and Discussion

The application of non-parametric statistical tests for the annual maximum daily flow series (Qd_{max}) is presented in Table 3, where we summarized the estimates of variables for the tests of randomness (run) and trend (SP, MK and PT).

Analysing the significance level in Table 2, when H_0 is 10%, it does not reject H_0 for $\alpha_0 = 10\%$. However, if H_0 is 5%, it does not reject H_0 for $\alpha_0 = 5\%$, and if it is H_a , it rejects H_0 regarding the trend.

For the significance level of 10%, the studied hydrological series was considered random (the statistic test Z_{Run} was -1.80). For the MK trend test, the calculated value of the Z_{MK} statistic was 0.98, which indicated non-rejection of the null hypothesis (i.e., the non-existence of a trend at a 5% significance level).

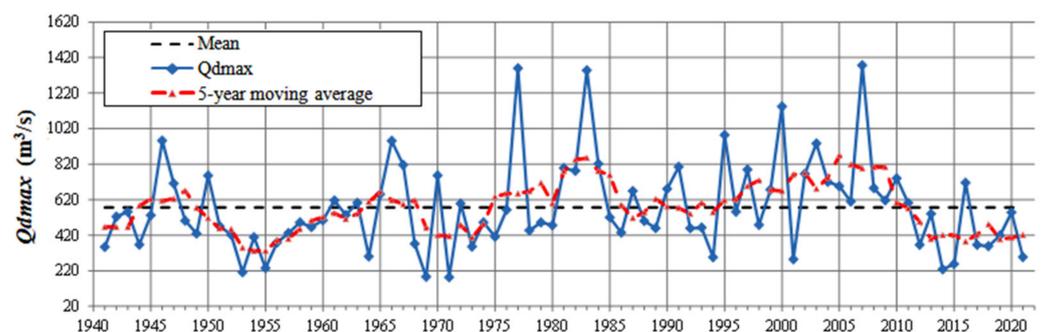
The module of the highest estimated value of the Pettitt test variable (Kt) was 459, which occurred for the year 1975. This result is lower than the critical limit ($k_{crit} = 576$), and therefore, the oscillations of the annual maximum daily water levels were stationary without the occurrence of any abrupt change in the series.

Table 3. Statistical non-parametrical tests of randomness and trend for the 4C-001 gauging station.

Run Test of Randomness	
Number of oscillations or runs	33
Statistic test : $Z_{run} = (r - E(u)) / \sqrt{Var(u)}$	-1.80
Significance level of the test	H0:10%
Trend	Random
MK trend test	
Test variable	1741
Statistic test : $Z_{MK} = (tn - E(tn)) / \sqrt{Var(tn)}$	0.98
Significance level of the test	H0:5%
Trend	NS
SP trend test	
Correlation coefficient	-0.11
Statistic test : $Z_{SP} = r / \sqrt{Var(r)}$	-0.98
Significance level of the test	H0:5%
Trend	NS ¹
PT trend test	
Critical level	1576
Statistic test : $Kt = MAX Ut $	459
Test meaning	No rupture in the hydrological series
Maximum occurrence year	1975
Minimum occurrence year	2011

¹ NS refers to no trend. When there is significant trend, it might be positive or negative.

We have elaborated the graph of the 4C-001 hydrological series to analyse the historical behaviour of the annual maximum daily water levels (Qd_{max}), as shown in Figure 2. The black and red dashed lines represent the multi-annual sample mean and the 5-year moving average, respectively.

**Figure 2.** Historical series of maximum annual flow for the gauging station 4C-001.

Throughout the centred five-year moving average of the maximum annual daily flow rate, three distinct periods for Qd_{max} in the Pardo River can be highlighted:

- From 1940 to 1975: the initial time interval of the historical series whose Qd_{max} values were either close to the sample mean, or lower. This period showed a descending behaviour of the hydrological variable.
- From 1975 to 2010: in this period of the hydrological series, the studied variable showed a strongly increasing behaviour, where the highest values of maximum flow of 1376.6 occurred in 2007, followed by 1348.8 in 1983. These higher magnitudes of Qd_{max} had not occurred in the series until previous years.

- From 2010: in the current period of the series, maximum flow values were either close to the sample mean or slightly below, presenting a descending behaviour with magnitudes close to those of the 1950 to 1960 interval.

In addition, Figure 3 presents the estimated Pettitt test statistic variable ($U(t,T)$) over the years. In this graph, it is observed that the minimum value of the test ($U(t,T) = -459$) occurred in 1975, whereas the highest value of the statistic occurred in 2011 ($U(t,T)=354$).

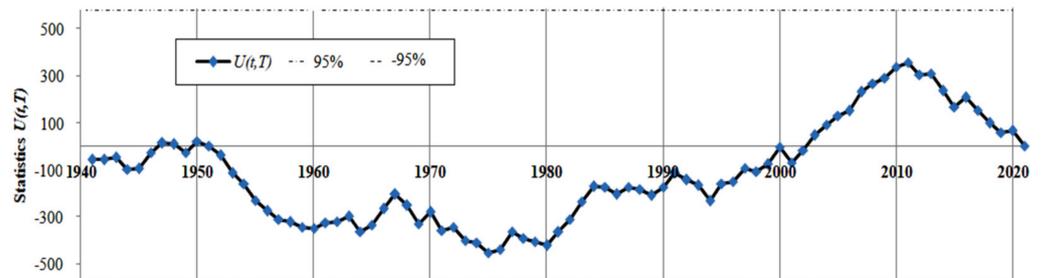


Figure 3. PT trend for the Qd_{max} series at the 4C-001 gauging station.

The period between 1975 and 2011 included the highest observed values of the annual maximum daily flow, whose behaviour of the moving average (Figure 2) was above the sample mean most of the time. During this time interval, the estimates of the Pettitt's statistical variable sharply increased. From 2012 on, it was observed that there was a trend for the flow to decrease. However, as this recent period corresponded to only 10 years, it was not possible to safely estimate either the sample mean or the standard deviation. The precipitation observed in the 2012–2021 period was lower than that observed in the previous period (1976–2012), but it was not long enough for the robust determination of its statistical parameters. Thus, this 10-year period was incorporated into the long wet period (1976–2021*), even knowing that this inclusion would mean a reduction in the statistics for the wet period.

The ascending period of the statistical variable of the PT reflected, at each new reference value (Y_i), how it would be higher than the previous reference value, describing the ascending trend of the sequential observations (long wet period). The number of data that were lower than the reference value was higher, until it reached the upper limit, which showed the end of the wettest period. After that point, there was a curve inflection, and the number of flows observed in the series became lower than the previous reference, marking the beginning of the descending stretch and corresponding to the long driest period. The interpretation of the statistical variable behaviour of the Pettitt Test is a useful tool for identifying and separating the long dry and wet periods of the hydrological series. Table 4 presents the results concerning the descriptive statistics of the maximum annual daily flow at different time intervals of the hydrological series. In addition to the sample mean and standard deviation (σ), the maximum and minimum observations are presented as well.

Table 4. Estimation of the mean, standard deviation, maximum and minimum observed values for the maximum annual daily flow (1941–2020).

Analysed Period	1941–2021	1941–1975	1975–2020
Number of years	81	35	46
Maximum observed annual daily flow (m^3/s)(year of occurrence)	1376.6 (2007)	952.7 (1946)	1376.6 (2007)
Minimum observed annual daily flow (m^3/s) (year of occurrence)	183.7 (1971)	183.7 (1971)	228.10 (2014)
Mean (m^3/s)	576.1	500.9	628.5
Standard deviation (m^3/s)	252.1	192.6	276.7

In Table 3, we observe a great variability of the annual maximum daily flows over time. For the same hydrological series, the mean and standard deviation of the sample are not constant over time, making it possible to compare the values of different wet periods (1975 to 2011) and dry periods (time intervals from 1941 to 1975).

For the continuous long wet periods, as represented by the years 1975 to 2011, it is observed that besides the elevation of the sample mean, there was a significant increase in the sample standard deviation. Conversely, in long dry periods, there was a reduction in both the mean and the standard deviation regarding the estimates of the long wet periods.

Compared with the Gumbel–Chow mathematical model, thoroughly explained by Naghettini and Pinto [32], the maximum probable flows could be estimated by the sum of the sample mean with the standard deviation multiplied by the frequency factor (k_t), which is a function of the recurrence interval ($Q_{dmax} = \mu + \sigma \times kt$).

During the long wet period (1975–2021*), both descriptive statistics presented higher values; therefore, according to the Gumbel–Chow mathematical model [32], it is expected that the estimates of maximum probable extreme flows are also higher when compared to the longest dry period [37].

In the central south region of Brazil, a reduction in annual rainfall was observed from the middle of the year between 1930 and 1970, which concurred with a lower predominance of El Nino events. This was followed by the period 1975–2012, when there was greater predominance of ENSO [38].

Analysing Figure 2, the maximum extremes occurred. During the long wet period, the highest value (1376.6 m³/s) was in 2007. However, in the long dry period, the lowest value (183.7 m³/s) was reported in 1971.

Zuffo and Zuffo [37] studied the historical series of annual precipitation of the City of Campinas, State of Sao Paulo from 1910 to 2014. They verified a great variability in precipitation over the years. The authors divided the hydrological series into four distinct plots to show that the mean and standard deviation of the sample were not constant over time. In the long dry periods of the study, there was a small drop in the average annual precipitation, but the major difference was in the standard deviation, which was much lower than those observed in the long wet periods, such as those represented by the years from 1910 to 1932 and from 1968 to 1991. Part of the study of Boulomytis et al. [12] was also about the long wet and dry periods of the northern coastline of the State of Sao Paulo, analysing the hydrological data between 1940 and 2015. In this study, there were fluctuations in the wavelengths of the two different patterns of cycles for about 32 years.

According to Zuffo and Zuffo [37] and Boulomytis et al. [12], the analysis period significantly influences the results of the adjustment of theoretical models of probability distribution for extreme events since the traditional methods for estimating the parameters do not incorporate the cyclical rise and recession features of long periods and neither the change in mean nor standard deviation, which are essential parameters for defining the stochastic model.

In the current study, we corrected the Q_{dmax} values from the long dry period (1941–1975) to make the hydrological series unique and homogeneous, according to the proposed methodology. In Figure 4, we observe the accumulated curves of the annual maximum daily flows as well as the respective linear regressions adjusted to remove cyclical trends and the final corrected curve.

After removing the temporal cyclical variability from the different periods, the parameters of the Gumbel probability distribution functions were adjusted to quantify the maximum probable flows as a function of the return period (Tr). Table 5 presents the estimates of the parameters of the theoretical Gumbel distribution function for the observed and corrected series. The corrected values of the mean, maximum and minimum observed are also updated.

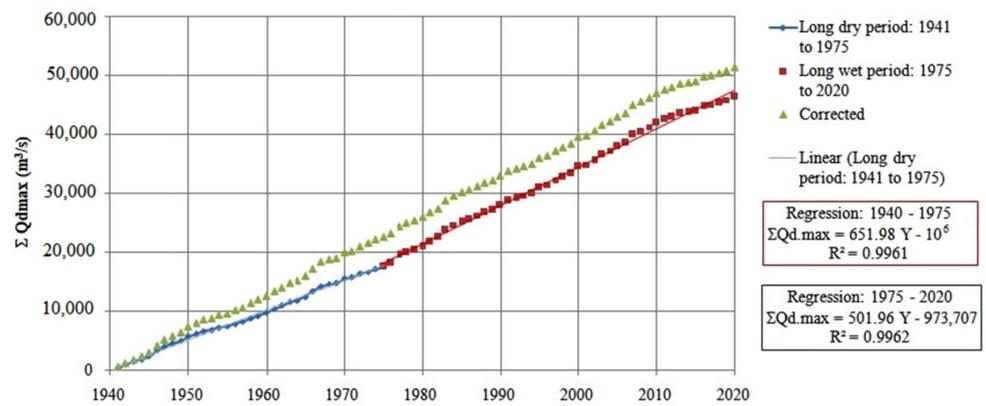


Figure 4. Correction for the cyclic variation of the long dry period (1941–1975).

Table 5. Estimated values regarding the maximum annual daily flow (1941–2020).

Hydrological Series	Observed	Corrected
Number of years	81	81
Maximum observed annual daily flow (m ³ /s) (year of occurrence)	1376.6 (2007)	1376.6 (2007)
Minimum observed annual daily flow (m ³ /s) (year of occurrence)	183.7 (1971)	228.1 (1971)
Mean (m ³ /s)	576.1	637.6
Standard deviation (m ³ /s)	252.1	252.1
Position parameter (α)	191.4	208.3
Scale parameter (β)	464.5	517.3
Adherence test results (KS)	Dist* = 4,1% * Dt = 17,8% ** H0:1% ***	Dist* = 13,8% * Dt = 17,8% ** H0:1% ***

Notes: (*) Maximum difference between theoretical Gumbel and empirical frequency estimates; (**) tabulated value of maximum difference of Kolmogorov–Smirnov test; (***) H0:1%—does not reject H0 hypothesis for α0 = 1% of Kolmogorov–Smirnov adherence test; Ha: rejects H0, and the theoretical model does not adhere to empirical frequency.

Analysing the results of the Kolmogorov–Smirnov adherence test for the hydrological series in Table 5, it is possible to observe that both are adjusted to the theoretical distribution function. In Figures 5 and 6, we present (a) the estimates of the partial frequencies and (b) the distance, or difference, between the theoretical and empirical probabilities for the observed and corrected historical hydrological series, respectively.

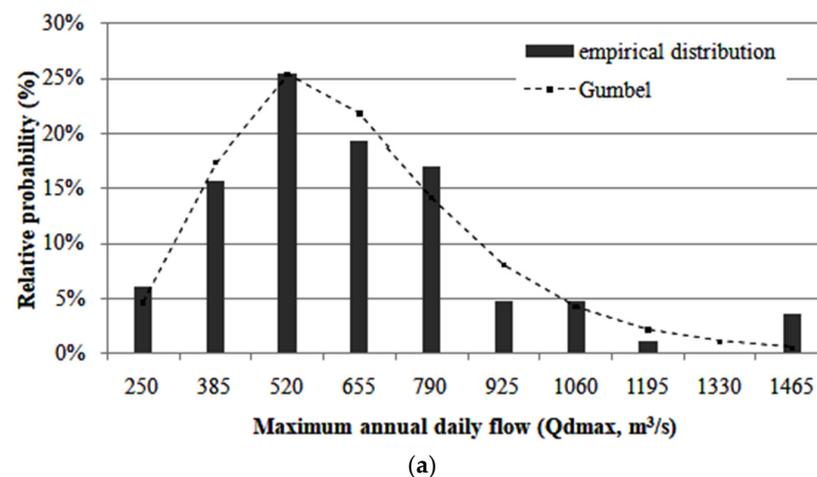
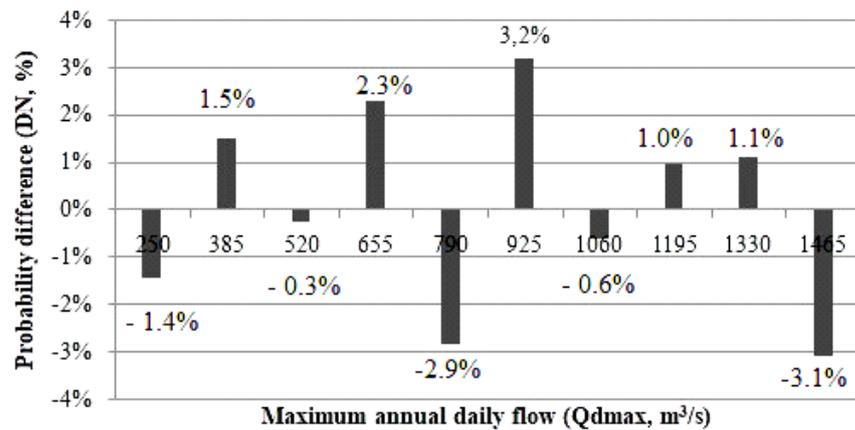
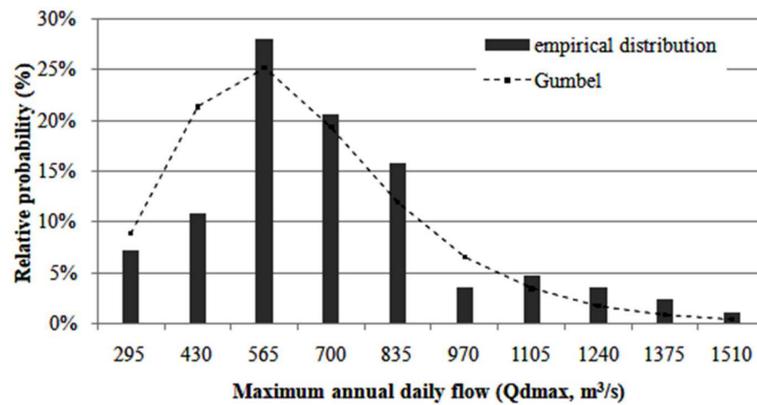


Figure 5. Cont.

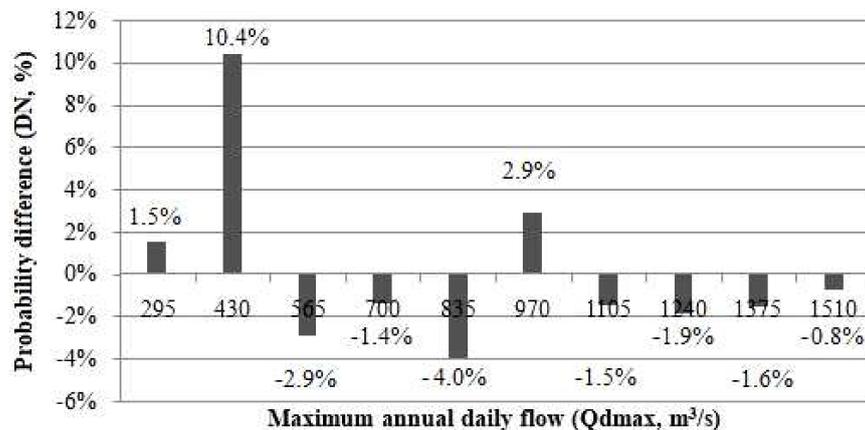


(b)

Figure 5. Estimated partial frequencies by Gumbel’s theoretical distribution model and empirical distribution for the observed hydrological series of maximum annual daily water flows from the gauging station 4C-001. (a) Partial frequencies estimated by Gumbel and empirical distribution functions. (b) Difference between Gumbel and empirical estimates.



(a)



(b)

Figure 6. Estimated partial frequencies by Gumbel’s theoretical distribution model and empirical distribution for the corrected hydrological series of maximum annual daily water flows from the gauging station 4C-001. (a) Partial frequencies estimated by Gumbel and empirical distribution functions. (b) Difference between Gumbel and empirical estimates.

From the parameters of the theoretical Gumbel distribution function (Table 5), the maximum probable annual daily flow rates were estimated for gauging station 4C-001 on the Pardo River. The results are presented in Figure 7, where we compare the estimated maximum probable flows for the observed and corrected historical series.

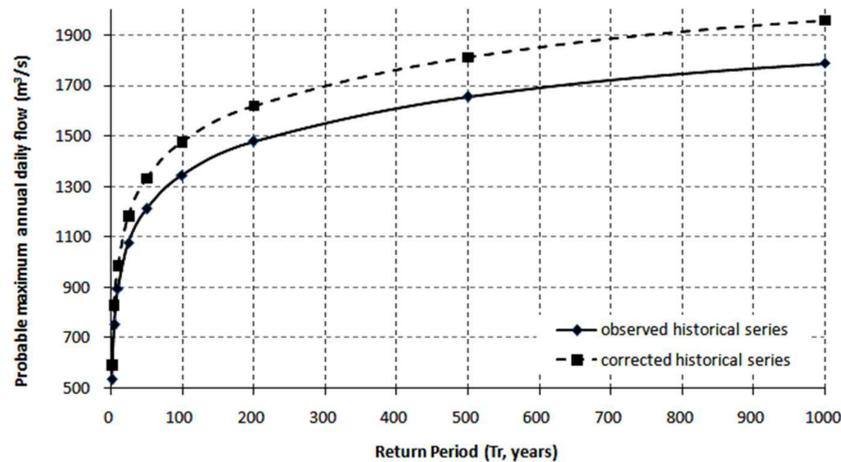


Figure 7. Probable maximum annual daily flows for gauging station 4C-001 on the Pardo River, from the observed and corrected hydrological series.

The application of the proposed correction in the long dry period made it possible to homogenize the magnitude of the sample statistics (mean and standard deviation). Thus, the estimates of probable maximum flows were around 10% higher for the corrected historical series when compared to the observed historical series. If these corrections are not made, the hydraulic structures designed from the observed historical series could be insufficient throughout the years of the long wet period. Since this behaviour of decadal variations presents a certain periodicity, it is probable that other periods similar to the long dry and the wet periods will occur. Therefore, when designing the hydraulic structures, which consider maximum extreme events, two alternatives are recommended:

- If the series is long enough and allows for separation into two subseries, one should comprise the long dry period and the other one, the long wet period.
- If the series is not long enough, adopt the correction of the long dry period for the corresponding wet period, as proposed in this study.

4. Conclusions

Through applying statistical tests, the historical series of maximum daily annual flows from the gauging station 4C-001 can be considered representative, where the variations around the mean are random, independent and without significant temporal trends. Although there are no trends in the historical series studied, it is possible to identify continuous periods where the maximum flows increase, describing long wet periods. These are followed by another long dry period and so on, which represent decadal variations. For these long dry and wet periods, when analysed separately, we observed significant differences in the estimation of the sample mean and standard deviation. In the long wet period, both mean and standard deviation are higher than those corresponding to the long dry period. The findings of the study showed that the proposed correction homogenized the sample statistics considering that the entire series would be a wetter period, and thus, the estimates of probable maximum flows found were around 10% higher for the corrected historical series if compared to the observed historical series. The difference in estimates could have been higher than the observed 10% values if the most recent dry period (2012–2021) had not been added to the long wettest period (1976–2021*). When the historical series is large enough in terms of number of years, and allowing for the identification

of more than one wetter long period spread with drier long periods, these series might be split into two parts, considering them as independent, which would improve the estimates of the extreme values. Otherwise, the methodology presented in this paper could be used. As the behaviour of long dry and wet periods is non-homogeneous, it is essential to distinguish their features for an accurate estimate of maximum or minimum extreme events, particularly when designing infrastructure assets (e.g., for the mitigation of floods and water supply reservoirs). This methodology promotes the estimate of more accurate extreme events and avoids the design of inefficient infrastructure assets.

Author Contributions: Conceptualization, M.A.J. and A.C.Z.; methodology, M.A.J.; writing—original draft preparation, M.A.J., A.C.Z. and V.T.G.B.; writing—review and editing, V.T.G.B., M.A.I. and A.C.Z.; visualization and final review, M.V.G. and T.A.S.; supervision, M.A.I. and A.C.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All the data is provided in the paper and cited references.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gitau, M.W. Patterns in indices of daily and seasonal rainfall extremes: Southeast Florida Gulf Coastal Zone. *Climate* **2018**, *6*, 83. [[CrossRef](#)]
2. Nerantzaki, S.; Papalexioiu, S.M. Assessing extreme in hydroclimatology: A review on probabilistic methods. *J. Hydrol.* **2022**, *605*, 127302. [[CrossRef](#)]
3. Hossain, I.; Imteaz, M.A.; Khastagir, A. Effects of estimation techniques on Generalised Extreme Value Distribution (GEVD) parameters and their spatio-temporal variations. *Stoch Environ. Res. Risk Assess.* **2021**, *35*, 2303–2312. [[CrossRef](#)]
4. Hossain, I.; Khastagir, A.; Aktar, M.N.; Imteaz, M.A.; Huda, D.; Rasel, H.M. Comparison of estimation techniques for generalised extreme value (GEV) distribution parameters: A case study with Tasmanian rainfall. *Int. J. Environ. Sci. Technol.* **2021**, *19*, 7737–7750. [[CrossRef](#)]
5. Yilmaz, A.G.; Imteaz, M.A.; Perera, B.J.C. Investigation of Non-stationarity of Extreme Rainfalls and Spatial Variability of Rainfall Intensity-Frequency-Duration Relationships: A case study of Victoria, Australia. *Int. J. Climatol.* **2017**, *37*, 430–442. [[CrossRef](#)]
6. Rezaei, K.; Vadiati, M. A comparative study of artificial intelligence models for predicting monthly river suspended sediment load. *J. Water Land Dev.* **2020**, *45*, 107–118. [[CrossRef](#)]
7. Ribeiro, L.U., Jr.; Zuffo, A.C.; da Silva, B.C. Development of a tool for hydroelectric reservoir operation with multiple uses considering effects of climate changes. Case study of Furnas HPP. *RBRH* **2016**, *21*, 300–313.
8. Asakereh, H. Decadal variation in precipitation regime in northwest of Iran. *Theor. Appl. Climatol.* **2020**, *139*, 461–471. [[CrossRef](#)]
9. Johnson, Z.F.; Chikamoto, Y.; Luo, J.-J.; Mochizuki, T. Ocean impacts on Australian interannual to decadal precipitation variability. *Climate* **2018**, *6*, 61. [[CrossRef](#)]
10. Li, S.; Gong, Z.; Zhang, S.; Yang, J.; Qiao, S.; Feng, G. Decadal variation of the precipitation relationship between June and August over South China and its mechanism. *Clim. Dyn.* **2022**, *58*, 1–20. [[CrossRef](#)]
11. Fusco, G.; Cotroneo, Y.; Alicino, G. Different behaviours of the Ross and Weddell Seas surface heat fluxes in the period 1972–2015. *Climate* **2018**, *6*, 17. [[CrossRef](#)]
12. Boulomytis, V.T.G.; Zuffo, A.C.; Imteaz, M.A. Derivation of Design Rainfall and Disaggregation Process of Areas with Limited Data and Extreme Climatic Variability. *Int. J. Environ. Res.* **2018**, *12*, 147–166. [[CrossRef](#)]
13. Kadioglu, M.; Ozturk, N.; Erdun, H.; Sen, Z. On the precipitation climatology of Turkey by harmonic analysis. *Int. J. Climatol.* **1999**, *19*, 1717–1728. [[CrossRef](#)]
14. Gericke, O.J.; du Plessis, J.A. Evaluation of critical storm duration rainfall estimates used in flood hydrology in South Africa. *Water SA* **2011**, *37*, 453–469. [[CrossRef](#)]
15. Arnbjerg-Nielsen, K.; Willems, P.; Olsson, J.; Beecham, S.; Pathirana, A.; Gregersen, I.B.; Madsen, H.; Nguyen, V.-T.-V. Impacts of climate change on rainfall extremes and urban drainage systems: A review. *Water Sci. Technol.* **2013**, *68*, 16–28. [[CrossRef](#)]
16. Alfieri, L.; Salamon, P.; Pappenberger, F.; Wetterhall, F.; Thielen, J. Operational early warning systems for water related hazards in Europe. *Environ. Sci. Policy* **2012**, *21*, 35–49. [[CrossRef](#)]
17. Youssef, A.M.; Maerz, N.H. Overview of some geological hazards in the Saudi Arabia. *Environ. Earth Sci.* **2013**, *70*, 3115–3130. [[CrossRef](#)]
18. Gaume, E.; Bain, V.; Bernardara, P.; Newinger, O.; Barbuc, M.; Bateman, A.; Blaškovičová, L.; Blöschl, G.; Borga, M.; Dumitrescu, A.; et al. A compilation of data on European flash floods. *J. Hydrol.* **2009**, *367*, 70–78. [[CrossRef](#)]

19. Fawad, M.; Cassalho, F.; Ren, J.; Chen, L.; Yan, T. State-of-Art Statistical Approaches for Estimating Floods Events. *Entropy* **2022**, *24*, 898. [CrossRef]
20. Pavlides, A.; Agou, V.D.; Hristoopoulos, D.T. Non-parametric kernel-based estimation and simulation of precipitation amount. *J. Hydrol.* **2022**, *212*, 127988. [CrossRef]
21. Totaro, V.; Gioia, A.; Iacobellis, V. Numerical investigation on the power of parametric and nonparametric tests for trend detection in annual maximum series. *Hydrol. Earth Syst. Sci.* **2020**, *24*, 473–488. [CrossRef]
22. Chow, V.T. *Handbook of Applied Hydrology: A Compendium of Water-Resources Technology*; McGraw-Hill: New York, NY, USA, 1964.
23. Koutsoyiannis, D.; Montanari, A. Negligent killing of scientific concepts: The stationarity case. *Hydrol. Sci. J.* **2015**, *60*, 1174–1183. [CrossRef]
24. Kendall, M.G. *Rank Correlation Methods*, 4th ed.; Charles Griffin: London, UK, 1975.
25. Lehmann, E.L. *Nonparametrics, Statistical Methods Based on Ranks*; Holden-Day: Oxford, UK, 1975.
26. Pettitt, A.N. A non-parametric approach to the change-point problem. *Appl. Stat.* **1979**, *28*, 126–135. [CrossRef]
27. Yue, S.; Pilon, P.; Phinney, B.; Cavadias, G. The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrol. Process.* **2002**, *16*, 1807–1829. [CrossRef]
28. Lall, U. Recent advances in nonparametric function estimation: Hydrologic applications. *Rev. Geophys.* **1995**, *33*, 1093–1102. [CrossRef]
29. Wang, W.; Van Gelder, P.H.; Vrijling, J.K. Trend and stationarity analysis for stream flow processes of rivers in Western Europe in the 20th Century. In Proceedings of the IWA International Conference on Water Economics, Statistics, and Finance, Rethymno, Greece, 8–10 July 2005.
30. Ribeiro Junior, L.U. Improvement of a Management Instrument for Operation of Hydropower Reservoirs with Multiple Uses under Conditions of Climate Change: Case Study Furnas Hydropower Plant. Ph.D. Thesis, State University of Campinas, Campinas, Brazil, 2013.
31. Pellegrino, G.Q. Spatial-Temporal Analysis of the Hydrological Data of Piracicaba River Watershed. Master's Thesis, University of Sao Paulo, Piracicaba, Brazil, 1995.
32. Naghettini, M.; de Andrade Pinto, E.J. *Hidrologia Estatística*; CPRM: Belo Horizonte, Brazil, 2007.
33. Back, A.J. Application of statistical analysis for identification of climatic tendency. *Pesqui. Agropecu. Bras.* **2001**, *36*, 717–726. Available online: <https://www.scielo.br/j/pab/a/YsG54Qqxr3msgZPRBD3nLvF/?format=pdf&lang=pt> (accessed on 25 April 2022). [CrossRef]
34. Detzel, D.H.M.; Bessa, M.R.; Vallejos, C.A.V.; Santos, A.B.; Thomsen, L.S.; Mine, M.R.M.; Bloot, M.L.; Estrochio, J.p. Estacionariedade das aflúncias às usinas hidrelétricas Brasileiras. *RBRH* **2011**, *16*, 95–111. [CrossRef]
35. Beijo, L.A. *Gumbel Distribution: Study of Methods of Estimating of Parameters and Adjustments to Data of Maximum Rainfall of Lavras*; UFLA: Lavras, Brazil, 2002.
36. Bautista, E.A.L. The Generalized Extreme Value Distribution to Study Maximum Wind Speed in Piracicaba, SP. Master's Thesis, University of Sao Paulo, Piracicaba, Brazil, 2002.
37. Zuffo, A.C.; Zuffo, M.S.R. *Water Resources Management: Conceptualization and Contextualization*; Elsevier: Rio de Janeiro, Brazil, 2016.
38. Mariani, M.; Fletcher, M.-S.; Holtz, A.; Nyman, P. ENSO controls interannual fire activity in southeast Australia. *Geophys. Res. Lett.* **2016**, *43*, 10891–10900. [CrossRef]