

Hurst exponent of hydrological time series obtained by R/S analysis and climacogram

Expoente de Hurst de séries hidrológicas obtido por análise R/S e climacograma

Mauricio Pereira Cantão
Marcelo Rodrigues Bessa
Daniel Henrique Marco Detzel
Victor Gabriel Monteiro

Programa de Pós-Graduação em Engenharia de Recursos Hídricos e Ambiental – PPGERHA
Universidade Federal do Paraná – UFPR
mpcantao@gmail.com
marcelo.bessa@ufpr.br
detzel@ufpr.br
victorgmonteiro@ufpr.br

Abstract:

Hurst exponent is a widely known long-term memory behaviour parameter for time series and it can be determined by several methods. In this work, two methods are analysed: the original R/S analysis and the more recent one, climacogram. Both are based on the scaling approach. Available tools for both methods are used in the analysis of the fractional Gaussian noise (fGn) series and the Brazilian inflow time series. The aim of analysing the fGn series is to establish the basis of comparison of both methods and to test possible improvements on the climacogram routine. Then, the methods are applied to annual inflow time series using a 148-hydropower plant database. The improvements on climacogram available routine result in different behaviours for the fGn and the inflow time series, according to the statistics applied to the results. Using georeferenced coordinates, a map of the Hurst exponent was made, indicating the higher persistence in the central region of Brazil. R/S analysis for the Paraná hydrographic region resulted in a decreasing Hurst exponent from the upstream to downstream direction, a behaviour which is not detected when using climacogram.

Keywords: Hydrological time series, Hurst exponent, R/S analysis, climacogram.

Resumo: O expoente de Hurst é amplamente conhecido como um parâmetro de memória de longo prazo de séries temporais que pode ser determinado com vários métodos. Neste trabalho, dois métodos foram analisados: o método original, análise R/S, e o mais recente, climacograma, ambos baseados na abordagem da análise de escala. Ferramentas disponíveis para ambos os métodos foram usadas na análise de séries de ruído gaussiano fracionário (fGn) e de vazões no Brasil. A análise de séries fGn teve o objetivo de servir de base de comparação do desempenho dos métodos e testar melhorias na rotina de cálculo do climacograma. Os métodos foram então aplicados a uma base de dados de 148 usinas hidrelétricas. As melhorias na rotina disponível apresentaram desempenho diferente nas séries fGn e de vazões, a partir da estatística dos resultados. Usando coordenadas georreferenciadas, elaborou-se um mapa do expoente de Hurst, indicando maior persistência na região central do Brasil. A análise R/S da região hidrográfica do Paraná resultou em expoente de Hurst decrescente a jusante, comportamento não identificado no uso do climacograma.

Palavras-Chave: Séries temporais hidrológicas, expoente de Hurst, análise R/S, climacograma.

1 Introduction

This work discusses the use of two mathematical tools for hydrological time series analysis, by means of the Hurst exponent determination using rescaled range (R/S) analysis and climacogram. After testing both methods in synthetic fractional Gaussian series, the analysis focused on annual inflow time series located in Brazilian hydrographic regions.

Hurst exponent is a long-range memory measure proposed by H. E. Hurst [1] during the calculation of River Nile Aswan Dam height. It relates with the persistence of inflow time series and its definition links up with the calculus procedure developed by Hurst, now called “rescaled range (R/S) analysis”. R/S analysis is based on the three ideal reservoir criteria [2] [3]: (i) uniform outflow, (ii) same water level at the beginning and at the end of an observation window, and (iii) no spills from the reservoir.

Hurst’s law is expressed by the relation in Equation 1 [4],

$$R_n/S_n = Cn^H \quad (1)$$

where R_n is the range of the first n cumulative deviations from the mean, and S_n is the sum of the first n standard deviations. Mandelbrot and Wallis [4] observed that Hurst used the relation $R_n/S_n \approx (n/2)^H$ instead of equation 1 (which, based on two parameters, allows for a better fitting) and asserted a different interval for H ([0.5, 1] instead of [0, 1]). O’Connell et al. [5] remembered that Hurst obtained $H = 0.5$ in simple coin toss experiment, while the Nile flow data resulted in $H = 0.729$ with standard deviation of 0.092. The difference between the values obtained by Hurst for hydrological time series (0.73) and with classical statistics (0.5) was called “Hurst phenomenon” [5].

Mandelbrot and Wallis [4] showed that the Hurst’s power law in R/S analysis relates to self-affine fractional Brownian motion (fBm) processes $B_H(t)$, establishing the basis for

fractal and multifractal time series analyses. The mathematics of $B_H(t)$ processes and their applications are discussed by Mandelbrot and Van Ness [6]. A native MATLAB® tool for fBm series synthesis was used and its performance is discussed with the help of $B_H(t)$ processes properties.

Several methods for Hurst exponent determination can be found in the literature, including power spectral density, autocorrelation function, detrended fluctuation analysis, wavelets, structure function etc. R/S analysis is the first method and serves as a basis for comparison.

Climacogram, or the standard deviation as a function of time scale [7], is a simple method for hydrological time series analysis. The method was proposed by D. Koutsoyiannis and his ITIA research team at the National Technical University of Athens as an analysis tool of the Hurst phenomenon.

The application of both methods to fBm processes and to hydrological annual time series showed that the climacogram algorithm developed by the Athens group can be improved with minor modifications.

The next section introduces the fBm process and discusses the wavelet tool used for generation of $B_H(t)$ series. The mathematical description of the techniques used in this work is also given in section 2, including its drawbacks, the meaning of results, and the software tools used in the calculations. A description of hydrological time series closes the section. Section 3 brings the results obtained with the analysed techniques. Conclusions appear in section 4.

2 Methods

This section comprises the computational tool used to synthesize fBm series, presents the R/S analysis and the climacogram methods for Hurst exponent determination, and explains the set of Brazilian hydrological time series analysed by means of the Hurst exponent.

Both R/S analysis and climacogram methods are second-order scaling approaches based on the standard deviation behaviour as function of the scale.

2.1 Fractional Brownian motion process

Mandelbrot and Ness introduced a family of Gaussian random functions and called them “fractional Brownian motion” (fBm) [6]. According to the authors,

“ $B(t)$ being ordinary Brownian motion, and H a parameter satisfying $0 < H < 1$, fBm of exponent H is a moving average of $dB(t)$, in which past increments of $B(t)$ are weighted by the kernel $(t - s)^{H-1/2}$.”

where t and s are different points of the fractional Brownian motion. Following Mandelbrot’s works, the properties of a normalized fBm process $B_H(t)$ are summarized as [8]:

- $B_H(t)$ has stationary increments
- $B_H(0) = 0$, and $\mathbb{E}[B_H(t)] = 0$ for $t \geq 0$
- $\mathbb{E}[B_H^2(t)] = t^{2H}$ for $t \geq 0$

- $B_H(t)$ has a Gaussian distribution for $t > 0$

$\mathbb{E}[\cdot]$ is the expected value, usually expressed by the mean. From the first three properties, it follows that the fBm covariance function is given by the expression in Equation 2 [8],

$$\mathbb{E}[B_H(s)B_H(t)] = [t^{2H} + s^{2H} - (t - s)^{2H}]/2 \quad (2)$$

where the exponent H is the same Hurst exponent proposed by H. E. Hurst. The non-stationary motion-like fBm process is associated with the “fractional Gaussian noise” (the fBm increment series, or fGn), which is stationary noise-like, as shown in Figure 1.

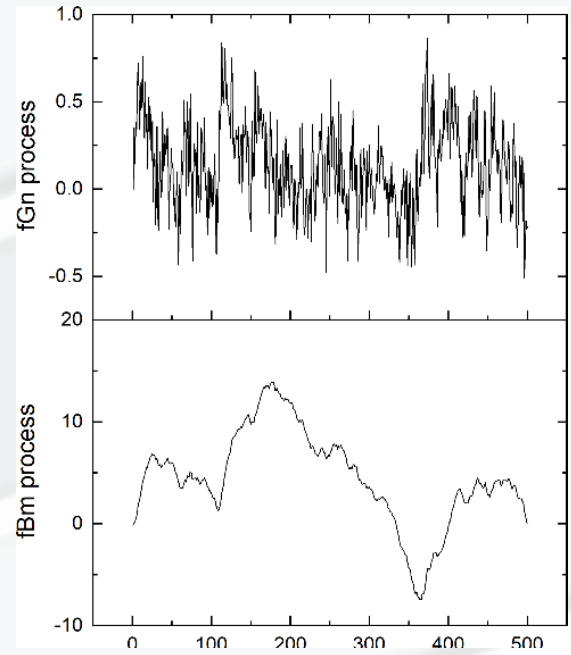


Figure 1: Noise-like fGn series (upper) and its associated walk-like fBm series (lower) for exponent $H = 0.8$.

The fBm processes show fractal properties, namely:

- Self-affinity: $B_H(at) \sim |a|^H B_H(t)$
- Long-range dependence
- Regularity by the fulfillment of Hölder condition
- Relation with fractal dimension, given by $D = 2 - H$

The symbol “ \sim ” means “equality in distribution”, D is the fractal dimension, e.g., Hausdorff or box-dimension, and “affinity” is used instead of “similarity” since ordinate and abscissa have different scaling factors, respectively $|a|^H$ and a . The interpretation of H depends on its range:

- $0.5 < H < 1$: Persistence. The series is long-term autocorrelated, i.e., a high value is probably followed by other high value and vice-versa.
- $H = 0.5$: Non-persistence. The series is completely uncorrelated.
- $0 < H < 0.5$: Anti-persistence. A high value in the series is probably followed by a low value and vice-versa.

For fBm series synthesis, the MATLAB® Wavelet Toolbox function $wfbm(H, N)$ has been used, where H is the rated Hurst exponent ($0.0 \leq H \leq 1.0$) and $N = 500$ is the fBm series length. The documentation for the $wfbm.m$ function [9] informs that, starting with the fBm process as an expression of fractional integral of white noise, the algorithm builds a biorthogonal wavelet from a given orthogonal one adapted to H exponent. Then, the sample path is reconstructed from the new wavelet using wavelet decomposition, independent coefficients from Gaussian realizations, and approximation coefficients from fractional ARIMA process [10].

This method tends to exhibit too many high-frequency components [9], a behaviour avoided using downsampling [11] by a factor δ . By opening the $wfbm.m$ function in the MATLAB® platform, it is possible to change the sampling parameter from the default value $\delta = 10$.

Evaluation of the effect of changing downsampling factor on the synthesis of fBm series was done using their property $\mathbb{E}[B_H^2(t)] = t^{2H}$, expressed as a logarithmic relation in Equation 3,

$$\log \mathbb{E}[B_H^2(t)] = 2H \log t \quad (3)$$

where H refers to the rated Hurst exponent. The expected value $\mathbb{E}[\cdot]$ means that Equation 3 applies to an *ensemble* of fBm processes and not to one time series, in which case the process would be *deterministic* and governed by the relation $B_H^2(t) = 2H \log t$. Equation 3 can be used as a method to assess the performance of tools for fBm time series synthesis. In this paper, it was used to analyse the effect of changing downsampling factor δ .

The following assessment procedure was used: for a set of 1,000 fBm series $B_H(t)$, the mean $\langle B_H^2(t) \rangle$ is calculated as a function of t . From the slope m of the plot $\log \langle B_H^2(t) \rangle \times \log t$, the expected Hurst exponent is $H_{\text{exp}} = m/2$. The process is repeated 100 times to verify the reproducibility and the dispersion of the simulation processes. Figure 2 shows the resulting log-log plots for different rated values of H for factor $\delta = 10$.

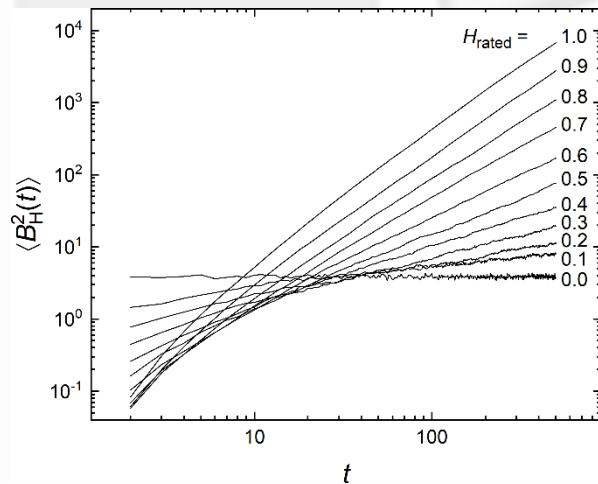


Figure 2: $\log \langle B_H^2(t) \rangle \times \log t$ plot used for performance assessment of the MATLAB® $wfbm.m$ function.

The procedure was repeated for different δ values (ranging from 1 to 200) for determining the optimal downsampling. Figure 3 demonstrates that each rated Hurst exponent has a factor δ which minimizes the difference $|H_{\text{exp}} - H_{\text{rated}}|$, where H_{exp} is the expected Hurst exponent obtained from equation 3. Even with the optimal δ the difference is higher for extreme Hurst exponents.

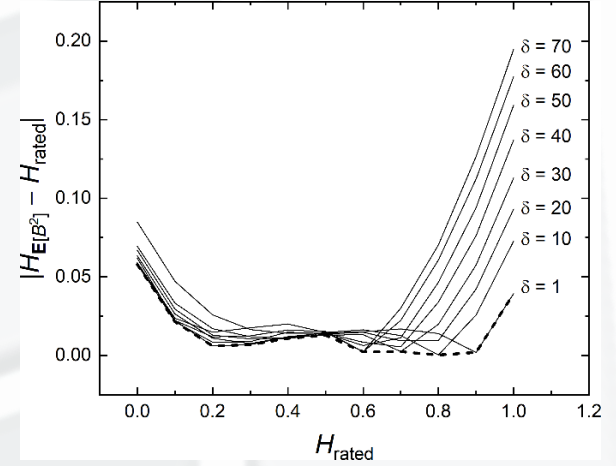


Figure 3: Absolute difference between H_{rated} and H_{exp} obtained using $\log \langle B_H^2(t) \rangle \times \log t$ plot (dashed line connects minimum values).

The effect of using an optimal (variable) δ against a default (constant) $\delta = 10$ is shown in Figure 4, without standard deviations. Changing δ results in Hurst exponents closer to rated values. The remaining deviation $|H_{\text{exp}} - H_{\text{rated}}|$ means that the $wfbm.m$ function is not able to generate fBm series with the exact (in statistical sense) H_{rated} for extreme values of the Hurst exponent.

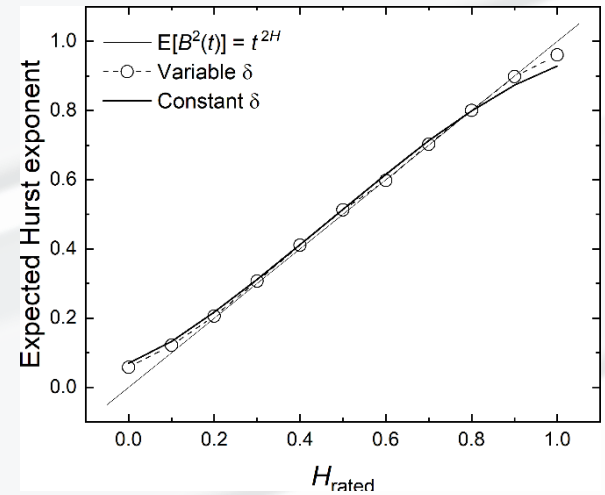


Figure 4: Expected Hurst exponent H_{exp} for different downsampling factors δ (H_{rated} given as a straight line).

Table 1 shows H_{rated} , the corresponding expected Hurst exponent H_{exp} , and the optimal downsampling factor. Admitting that the difference $|H_{\text{exp}} - H_{\text{rated}}|$ arises from a limitation of $wfbm.m$, the methods for the Hurst exponent determination for fGn series should be compared with the expected values in Table 1. The synthetic fBm series are

obtained for each of 11 H_{rated} values using the optimal downsampling factors of Figure 3 and Table 1. Again, the procedure starts with sets of 1,000 fBm series. The fGn series are obtained by simple differentiation using MATLAB® diff.m function and inserting a zero value as the first value. In this way, both fBm and the corresponding fGn series comply with the rule $B_H(0) = 0$ and exhibit the same length.

Table 1: Expected Hurst exponents, deviations from rated H and downsampling factors for optimal fBm synthesis.

H_{rated}	H_{exp}	Downs. factor δ
0.0	0.058 ± 0.004	60
0.1	0.121 ± 0.005	50
0.2	0.206 ± 0.007	40
0.3	0.307 ± 0.007	40
0.4	0.411 ± 0.008	20
0.5	0.513 ± 0.007	30
0.6	0.598 ± 0.010	70
0.7	0.702 ± 0.008	30
0.8	0.800 ± 0.009	10
0.9	0.898 ± 0.008	1
1.0	0.961 ± 0.007	1

The following subsections describe R/S analysis and climacograms for the Hurst exponent determination.

2.2 R/S analysis

R/S analysis is done by means of Equations 4 and 5 [2],

$$\left(\frac{R}{S}\right)_n = \frac{1}{S} \left[\max_{1 \leq k \leq n} \sum_{i=1}^k (x_i - \bar{x}) - \min_{1 \leq k \leq n} \sum_{i=1}^k (x_i - \bar{x}) \right] \quad (4)$$

$$\lim_{n \rightarrow \infty} \left(\frac{R}{S}\right)_n = Cn^H \quad (5)$$

where $k \in n$. For the observation window n , $R(n)$ is the range of accumulated values, $\bar{x} = \langle x \rangle_n$ is the average of time series $x(t)$, $S(n)$ is the standard deviation, and C is a constant. Often called “subseries length” [12] [13], the window n implies the scale approach of R/S analysis: portions of the time series with different lengths reflecting the whole behaviour of the underlying process. $R(n)$ coming from a summed series means that the R/S method is adequate for fGn noise-like series (see Figure 1). $H_{R/S}$ results from the plot of $\log R/S \times \log n$.

Calculation of $H_{R/S}$ involved the use of the OriginPro® software tool [12]. The chosen version is corrected for small sample bias and is based on the work from R. Weron [13]. Expressions for expected values $\mathbb{E}(R/S)_n$ for $n \leq 340$ and $n > 340$ can be found in Weron’s paper.

The OriginPro® app is adequate for small sets of time series, as is the case of hydrological time series. For larger sets of fGn series the MATLAB™ based function `hurst.m` [14] was chosen.

2.3 Climacogram analysis

For a statistically independent sample of n observations, the standard error is defined as $\sigma_{\bar{X}} = \text{std}[\bar{X}] = \sigma/\sqrt{n}$, where σ is the standard deviation of the population, and \bar{X} is the sample

mean. Koutsoyiannis [15] observed that the Hurst phenomenon can be expressed by the modified expression represented in Equation 6,

$$\text{std}[\bar{X}_n] = \sigma/n^{1-H} \quad (6)$$

where H is the Hurst exponent. For a time series X_i with a sufficient length N , the standard deviation $\sigma_k = \text{std}[\bar{X}_k]$ is defined by the sample means $\bar{X}_k^{(n)}$, where n is the sample length, usually a fraction of N . For example, if $k = 2$, the samples are: $\bar{X}_2^{(1)} = (X_1 + X_2)/2$, $\bar{X}_2^{(2)} = (X_3 + X_4)/2$, and so on [15]. The plot $\log \text{std}[\bar{X}_k] \times \log k$ is the climacogram of time series X_i . Scaling characteristic of the climacogram is related to the sample length k . Together with R/S analysis, the climacogram method is a scaling approach for time series analysis, in opposition to the classical statistics [5].

It is worth noting that Koutsoyiannis has proposed the scaling approach as the antipode of nonstationarity approach [15], in the sense that the last one demands the separation of the time series in components (e.g., trend), while the scaling analysis models the trend as a large-scale stochastic component. In this way, the author states that “the existence of trends is the normal behaviour of real-world time series” [15], implying that the non-stationary behaviour is part of a random process in sufficient large time scale.

The scope of this work does not aim at joining the scaling/stationarity vs. classical statistics/non-stationarity debate. Koutsoyiannis and his group admit that non-stationarity could be accepted if it is associated with a clear physical explanation [5]. It suffices to say that climacogram is a valid technique for time series characterization and that is the focus of this work.

The MATLAB® based function `Climacogram.m`, developed by G. Pouliasis for the calculation of the \bar{X} samples and their respective standard deviations, is available on the GitHub platform [16]. The function does not include bias correction (which is discussed in related works [15] [17] [18]) and an empirical rule states that the number of samples is given by 10% of the series length ($n = N/10$). The slope m of the $\log \sigma_k \times \log k$ plot (climacogram) relates to the Hurst exponent through the relation $H = 1 - |m|$, as seen in equation 6.

The Pouliasis’ routine has two features worth considering. First, for non-integer values of N/k the samples $\bar{X}_k^{(n)}$ do not cover the entire series X_i . If, for instance for $N = 90$ (the length of annual inflow series), the ratio N/k is integer for $k = 1, 2, 3, 5, 6, 9$. For $k = 4$ and 8 the value $k \cdot \text{int}(N/k) = 88$ and, for $k = 7$, the result is 84, which means that 2 or 6 values of the series stay out of the climacogram.

One way of addressing the entire time series is to take the series and analyse it as given (“original” climacogram) and after inversion (“inverted” climacogram). Figure 5 shows that the climacogram of the inverted series is unaffected when N/k is an integer, showing different results for the non-integer N/k values.

As an alternative to the conventional procedure, it is proposed to take the averaged climacograms of both

“original” and “inverted” climacograms as the “corrected” climacogram, also shown in Figure 5.

As an example of the effect of inverting the time series, one could the Nile River case, analysed by Koutsoyiannis [15]. Table 2 shows the results for the original, inverted, and corrected climacograms for the 663-year annual inflow time series [19]. Koutsoyiannis found $H = 0.85$ for the same Nilometer series.

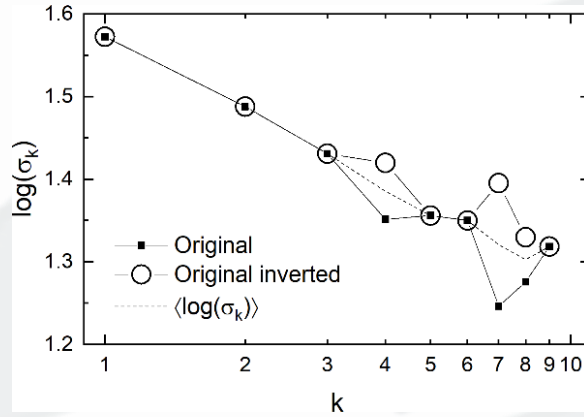


Figure 5: Climacograms for a time series with $N = 90$ and $k = 1, \dots, 9$, taken for original and inverted time series (average climacogram given as dashed line).

Table 2: Climacogram analysis for the Nile River. “Corrected⁽¹⁾” refers to the mean between “original” and “inverted”.

Result	Original	Inverted	Corrected ⁽¹⁾
H	0.858	0.853	0.855
σ_H	0.004	0.005	0.003
R^2	0.9395	0.9312	0.9726

As expected, the Hurst exponent for the corrected climacogram is the average between original and inverted climacograms, but the standard deviation σ_H is lower than the other two. Besides, the coefficient of determination R^2 for the linear regression is about 4% higher for the corrected climacogram, which indicates the benefit of this correction procedure.

A second issue comes from the compression of points of large-scale values of the linearly distributed abscissa k in a log-log plot, as shown in Figure 5. Eke et al. [3], for example, pointed out that determining the Hurst exponent with power spectral density suffers from overweighing of high frequencies, which should be avoided in the calculation. Climacogram plots resemble power spectral density, if σ_k is changed by the power spectral density and the scale k by the frequency (Koutsoyiannis calls the climacogram “pseudospectrum” [20]). Since H is determined through linear regression of the climacogram $\log \sigma_k \times \log k$ the gathering of points increases the weight of the larger scales, affecting the result.

The solution proposed in this work was adapted from the routine by Wen and Liu for power spectral density evaluation [21] [22]. It starts with a $\log k$ scale linearized with the operation $\log k' = \text{linspace}(\log k_1, \log k_2, n)$,¹ where $\log k_1$

and $\log k_2$ are the extremes of $\log k$, n is the sampling length and `linspace.m` is a MATLAB® function. Using the linear interpolation function `interp1.m` an alternate version $\log \sigma'_k \times \log k'$ is obtained from the original climacogram.

Figure 6 shows the same original climacogram from Figure 5 and its alternate linearly spaced scale version.

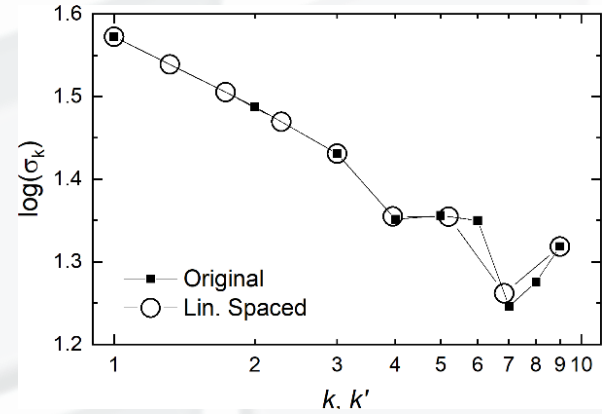


Figure 6: Climacograms for a time series with $N = 90$, taken for original (k) and linearly spaced (k') scales.

Table 3 shows the correspondent results obtained for Nilometer time series. It is seen that the linear sampling has a greater benefit than the inversion, reducing standard deviation and increasing R^2 . As some points are still missing after linear sampling, the inversion procedure was also tried, with no change of $H \pm \sigma_H$ and a reduction of R^2 . Hurst exponent determined with both corrected climacograms are close to the second decimal place.

Table 3: Climacogram analysis for Nile River. “LS”: linearly spaced; “LS/I”: inverted after linearly spaced; “Corrected⁽²⁾”: mean between “LS” and “LS/I”.

Result	LS	LS/I	Corrected ⁽²⁾
H	0.850	0.850	0.850
σ_H	0.002	0.002	0.002
R^2	0.9906	0.9835	0.9923

Other modifications of the original Pouliaxis’ function Climacogram.m include the insertion of the routine in a loop, allowing the automatic calculation of all the series in each set, and the use of the polynomial curve fitting `polyfit.m` instead of `fit.m`, as the first is a basic MATLAB® function while the second one is part of specific toolboxes.

2.4 Hydrological time series

The Brazilian National System Operator (ONS) provides monthly inflow time series from Jan/1931 up to Dec/2020. The inflows are the difference between affluence and effluence (incremental inflow), corrected by the consumptive uses, evaporation, and hydropower plants (HPP) operation [23]. After this reconstitution procedure, they are referred to as “natural inflows”.

Reconstituted natural inflows carry the uncertainty associated with the natural/artificial processes removed

¹ Linearly spaced values generator.

from the measured data. The reason for their use in hydrological studies is that they represent the behaviour of the natural resource, while measured inflows are affected by human activities. An example of natural inflow times series is shown in Figure 7 (Marimbondo HPP), together with the corresponding annual average of monthly inflows.

This example evidences the seasonality of monthly basis inflow. Since the scope of this work does not include this phenomenon, the analysis will focus on the annual inflow data.

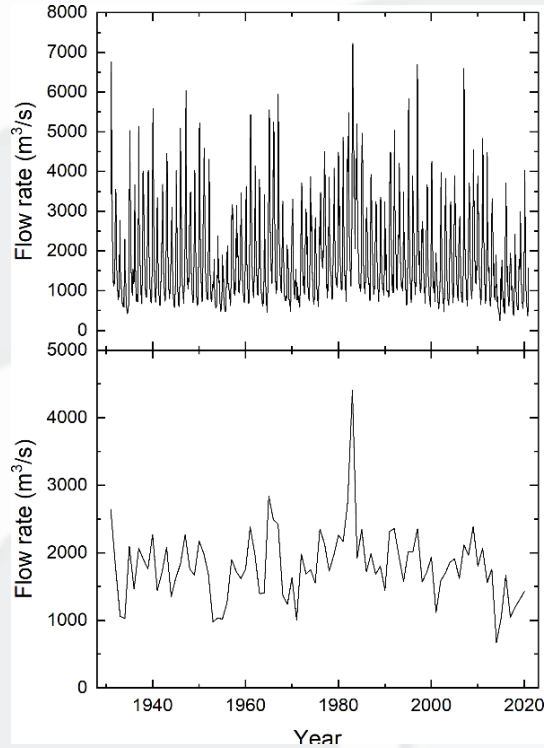


Figure 7: Inflow for Marimbondo HPP (Paraná Basin) in monthly (*top*) and annual (*bottom*) basis.

ONS inflow database is composed of 217 time series, including fictitious power plants and other series that do not represent water resources engineering works with natural affluence, such as transposition between reservoirs and pumping power plants. Removing these events and considering only operating power plants with natural affluence, the database reduces to 148 hydropower plants, HPP. The criteria of analysis are the hydrographic region, HPP type (run-of-the-river power plants, reservoir power plant with regulation, and plain reservoir), and whether the HPP is located as the most upstream reservoir or not.

3 Results

This section shows the results for fractional Brownian motion and hydrological time series.

3.1 Fractional processes

Figure 8 resumes the analysis of the fGn series with R/S and climacogram methods. The Hurst exponents $H_{R/S}$ and H_{clim} show similar behaviour in comparison with the expected

Hurst exponent H_{exp} (equation 3). $H_{R/S}$ is closer to H_{exp} for higher H_{rated} , but show large deviation when H_{rated} is small. It can be seen that $H_{clim} \approx H_{exp}$ for H_{rated} lower than 0.7 and that its standard deviation is lower in comparison to $H_{R/S}$. All considered, it can be said that the climacogram shows better performance than the R/S analysis of fGn series.

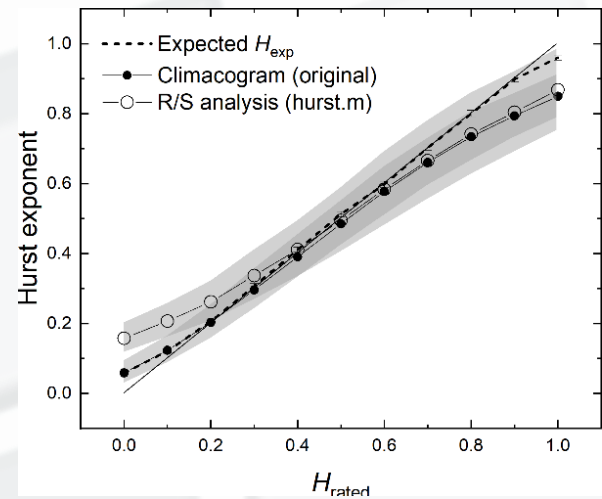


Figure 8: Average Hurst exponents for 10 rated H obtained with R/S and climacogram analyses (H_{rated} : straight line).

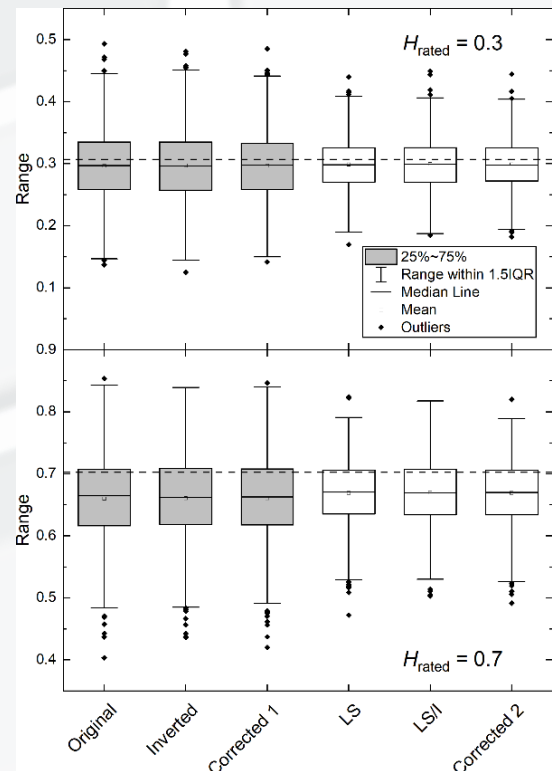


Figure 9: Boxplot statistics for two rated Hurst exponents (grey shaded boxplots refer to the original and the inverted fGn series and white boxplots relate to the linearly spaced scale k').

The statistical analysis of the Hurst exponent determined with climacograms was made with boxplots, shown in Figure 9 for $H_{Rated} = 0.3$ and 0.7. Results for the original fGn series

are given as grey shaded boxes, while the white boxes represent climacograms obtained with the linearly spaced scale k' . In each case, the inverted series, and the mean of inverted and non-inverted series ("Corrected" climacograms) are also shown. H_{exp} is represented as a straight dashed line in each plot.

Figure 9 demonstrates, for fGn series, that the inversion of the series does not change the statistical behaviour. Linear sampling of $\log k'$, on the contrary, reduces the range of the first and the third quartiles, the 1.5 interquartile (whiskers) range, the difference between average H_{clim} and H_{exp} , and the number of outliers. The inversion of fGn series after performing linear sampling did not improve the statistics of H_{clim} .

3.2 Hydrological time series – Brazil

Small sample corrected $H_{R/S}$ for the 148 annual inflow time series is 0.58 ± 0.06 . Arbitrating that $H = 0.50 \pm 0.02$ is the range of uncorrelated non-persistence, it was found that only 27 time series are non-persistent, and 121 HPP show persistent behaviour. For H_{clim} , only six HPP have non-persistent time series under the same arbitrary range.

The analysis per hydrographic region is shown in Table 4. It is observed that $H_{clim} > H_{R/S}$ in all cases, the difference ranging from 14% (Southern Atlantic) up to 26% (Paraguay). Considering all 148 HPP, only four time series (located in the Southern Atlantic region) show $H_{clim} < H_{R/S}$.

Table 4: Hurst exponents for the different hydrographic regions (148 HPP inflows).

Hydrographic region	$H_{R/S}$	H_{clim} (linearly spaced)
Amazon	0.59 ± 0.08	0.68 ± 0.12
Tocantins-Araguaia	0.64 ± 0.02	0.76 ± 0.04
Parnaíba	0.67	0.80
São Francisco	0.68 ± 0.03	0.81 ± 0.04
Eastern Atlantic	0.60 ± 0.06	0.72 ± 0.08
Paraguay	0.68 ± 0.03	0.86 ± 0.06
Paraná	0.57 ± 0.05	0.70 ± 0.07
Southeastern Atlantic	0.59 ± 0.05	0.69 ± 0.08
Southern Atlantic	0.55 ± 0.02	0.61 ± 0.10
Uruguay	0.52 ± 0.01	0.59 ± 0.04

Inflow time series at Uruguay hydrographic region show the lowest persistence as determined by both methods. The ten regions can be grouped by average Hurst exponent values, with little difference in the order. Taking as threshold values $\bar{H}_{R/S} = 0.60$ and $\bar{H}_{clim} = 0.72$, six regions (Amazon, Eastern Atlantic, Paraná, South-eastern Atlantic, Southern Atlantic, and Uruguay) have Hurst exponents computed from both methods below the thresholds $\bar{H} \leq \bar{H}$. These six regions include 128 of the 148 river most upstream inflow time series, with an average inflow of $1,024 \text{ m}^3/\text{s}$. Four hydrographic regions (Tocantins-Araguaia, Parnaíba, São Francisco, and Paraguay) show $H > \bar{H}$. This group is composed by the remaining 20 inflow time series, with average inflow equal to $1,676 \text{ m}^3/\text{s}$.

Figure 10 shows the boxplot analysis for the climacograms obtained for the 148 inflow time series provided by ONS for operating the HPP. The average \bar{H}_{clim} changes from 0.688 (for original time series) to 0.711 (inverted linearly sampled time series), an increase of 3.3%. Box/whiskers ranges also increase due to the inversion of the time series together with outlier removal. In this way, the mean between linearly spaced climacograms of the original and inverted ("Corrected 2") series offers a more reliable Hurst exponent H_{clim} for the inflow time series.

Figure 11 shows the linearly spaced corrected H_{clim} for the set of the 148 HPP located in Brazilian territory. Since there are no hydropower plants in the two North-eastern Atlantic regions and, the Parnaíba region has only one HPP, these regions were merged with São Francisco hydrographic region.

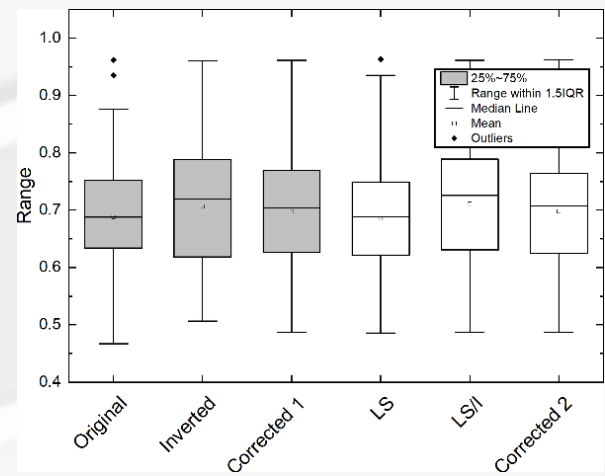


Figure 10: Boxplot statistics for 148 inflows (grey shaded boxplots: original and inverted time series, white boxplots: linearly spaced scale k').

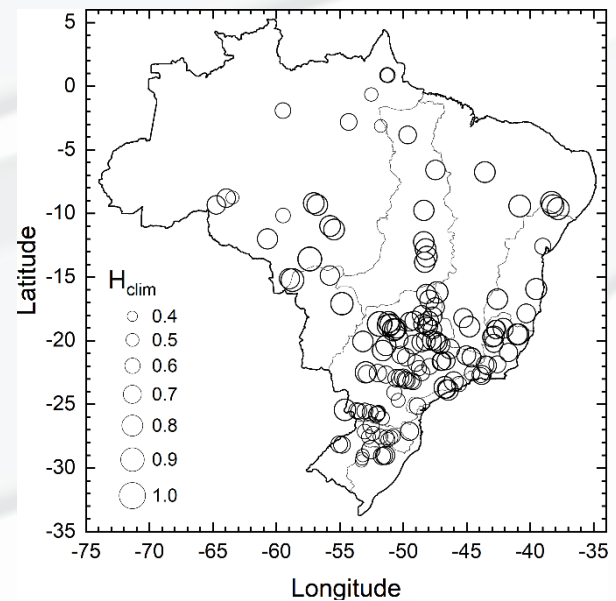


Figure 11: Brazilian map showing the location of the 148 operating HPP (corrected H_{clim} is shown in bubble plot and the hydrographic regions limits in grey).

Only four HPP showed $H_{\text{clim}} \leq 0.5$: Dona Francisca, Passo Real, and Jacuí (located in the Southern Atlantic hydrographic region), and Samuel (at Amazon region). Persistence ($H_{\text{clim}} > 0.5$) varied along the territory, with predominance of higher values in the Centre-Western and Northeastern regions (as seen in Table 4). The Amazon region has no border with the other hydrographic regions with $\bar{H} \leq \bar{H}$, indicating that the Hurst exponent behaviour is not controlled by geographic location. Average Hurst exponents with both methods are $\bar{H}_{R/S} = 0.58 \pm 0.06$ and $\bar{H}_{\text{clim}} = 0.70 \pm 0.10$ for the set of the 148 HPPs.

Restricting the Hurst exponent analysis to power plants located at the most upstream river locations avoids ambiguities with cascading power plants data (in general, obtained by linear regression). Since the calculation of inflow time series involves removing the effect of HPP operation, the most upstream river located power plants and reservoirs are less affected by this procedure. For the 50 river the most upstream points (reservoirs, run-of-the-river HPPs, and HPPs with regulated reservoirs) the analysis results are $\bar{H}_{R/S} = 0.59 \pm 0.07$, and $\bar{H}_{\text{clim}} = 0.70 \pm 0.11$, higher but still very close to the values obtained for all 148 HPPs.

The analysis of inflows according to the kind of HPP together with the most upstream river effect is shown in Table 5. The most upstream river HPP show higher Hurst exponents than the 148-HPP database for run-of-the-river HPPs and regulated reservoirs. The most upstream river reservoir HPP has the same values of H , but present higher dispersion.

The HPP type also affects the performance of the Hurst exponent methods. Considering the 148-HPP (river head) databases, the difference between H_{clim} and $H_{R/S}$ is 21% (21%) for the run-of-the-river HPPs, 18% (18%) for regulated reservoir HPPs, and 23% (37%) for plain reservoirs.

Table 5: Hurst exponents for the different hydropower plants.

HPP type (148 HPP)	$H_{R/S}$	H_{clim} (Corr. 2)
Run-of-river HPP (81 pts.)	0.59 ± 0.06	0.71 ± 0.10
Dam HPP (61 pts.)	0.58 ± 0.06	0.69 ± 0.10
Reservoir (6 pts.)	0.56 ± 0.04	0.67 ± 0.09

HPP type (river head, 50 HPP)	$H_{R/S}$	H_{clim} (Corr. 2)
Run-of-river HPP (18 pts.)	0.61 ± 0.07	0.72 ± 0.11
Dam HPP (31 pts.)	0.58 ± 0.07	0.69 ± 0.11
Reservoir (1 pt.)	0.53	0.71

One last comparison of both methods for the Hurst exponent determination is the linear Pearson correlation as a function of the hydrographic region. Table 6 shows the correlation coefficient between the sets of H values obtained with both methods for all HPPs of each region (descending order of r_p). With only one HPP, Parnaíba region is not on the list.

No geographic pattern is identified. For example, the group with higher correlation coefficients ($r_p \geq 0.9$) includes hydrographic regions in the North, Northeast, and Centre-West of Brazil. There is a second group with $0.6 < r_p < 0.9$ located in the Centre, Southeast, and South of Brazil. It is noted that the two hydrographic regions with lower

correlation coefficients are situated in the Brazilian Southern and Northeastern geographic regions, the Uruguay region being the only one with no correlation.

Table 6: Pearson correlation coefficient r_p between Hurst exponent methods.

Hydrographic region	Number of HPP	$r_p[H_{R/S}, H_{\text{clim}}]$
Eastern Atlantic	4	0.984
Amazon	17	0.951
Tocantins-Araguaia	7	0.906
Paraná	67	0.814
Southeastern Atlantic	18	0.731
Paraguay	5	0.658
Southern Atlantic	10	0.642
São Francisco	7	0.449
Uruguay	12	0.193

Coefficients of correlation higher than 0.8 (strong correlation) can be accepted as statistical differences in the methods performance. The behaviour of the Uruguay (no correlation) and São Francisco regions (weak correlation) indicates that both inflow time series are more sensitive to the Hurst exponent method.

Besides the separation of hydrographic regions according to the Hurst exponents below or above \bar{H} ranges and a trend of higher H at the centre of the Brazilian territory, no pattern can be seen from Figure 11. In this way, the Paraná hydrographic region (with almost half of the HPPs) is worth analysing.

3.3 Hydrological time series – Paraná Region

The 67 HPPs of the Paraná hydrographic region are shown in Figure 12. The five river basins that integrate the region are (from upstream to downstream): Parnaíba, Grande, Tietê, Paranapanema, and Iguaçu. Besides the river basins there are the HPPs located in the Paraná River itself: Ilha Solteira, Jupia, Porto Primavera, and Itaipu.

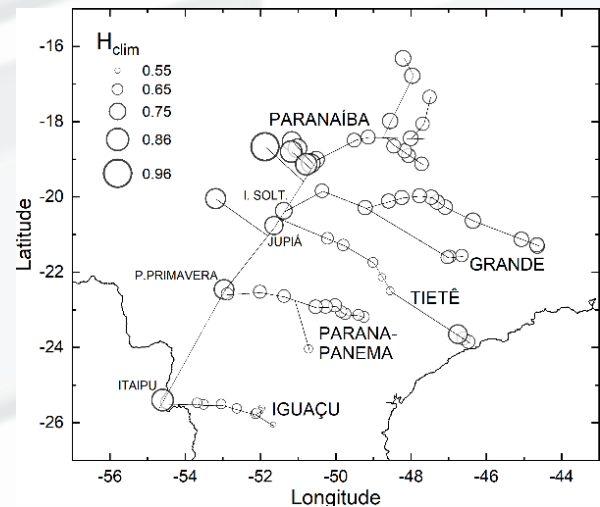


Figure 12. Bubble plot of Paraná hydrographic region and its river basins.

Espora HPP shows the highest value for the set of the Hurst exponents ($H_{\text{clim}} = 0.96$). This and other five power plants are referred to as “Paranaíba subgroup”, as they are closer to each other (just upstream Ilha Solteira HPP) and have higher H_{clim} than the rest of the basin’s HPPs. The lowest H_{clim} occurs for Foz do Areia, the first HPP in the Iguaçu basin.

Table 7 shows the results of R/S analysis and the climacogram for the HPPs in the Paraná hydrographic region.

Table 7: Hurst exponents for the Paraná river basins.

River basin	$H_{R/S}$	H_{clim} (inverted)
Paranaíba	0.62 ± 0.03	0.76 ± 0.07
Paranaíba subgroup	0.65 ± 0.02	0.85 ± 0.06
Grande	0.57 ± 0.01	0.72 ± 0.01
Ilha Solteira HPP	0.57	0.76
Tietê	0.57 ± 0.04	0.71 ± 0.08
Jupia HPP	0.58	0.79
Porto Primavera HPP	0.58	0.82
Paranapanema	0.53 ± 0.01	0.67 ± 0.06
Itaipu HPP	0.54	0.84
Iguaçu	0.50 ± 0.01	0.60 ± 0.03

The Paranaíba subgroup and the four Paraná River HPPs show the highest Hurst exponents obtained with climacogram ($H_{\text{clim}} \sim 0.8$); when using R/S analysis, only the complete Paranaíba basin shows $H_{R/S} > 0.6$. Table 7 indicates that $H_{R/S}$ decreases the further downstream the HPP is, an undetected behaviour for H_{clim} .

4 Conclusions

In this work, two methods for Hurst exponent calculation (R/S analysis and climacograms) are evaluated. R/S analysis is performed with two computing tools (MATLAB® and OriginPro®), both using a correction for small number of observations. Climacograms were obtained with a routine based on work by ITIA research group, modified in two fashions: the inversion of time series and the linear sampling on log-scale.

A MATLAB® function was used for fractional Gaussian noise (fGn) synthesis, optimized with downsampling factor tuning. Climacograms of fGn series showed that inverting the time series did not change the performance (measured as the absolute difference against the rated Hurst exponent), but linear sampling improved the statistics, by means of reduction of quartile ranges and number of outliers. The reason for the resulting effect of linear sampling of $\log k'$ resides on the homogeneous weighting through the scale during the linear regression of the climacogram $\log \sigma'_{k'} \times \log k'$.

The available routine for climacogram calculation was improved with two modifications: the inversion of time series, which avoids the exclusion of some points from the analysis, and the linear sampling of the logarithmic scale, which balances the weight of all scales in the final climacogram.

The effect of linear sampling on the annual inflow time series is not as evident as for fGn series, probably due to the series length ($N = 90$), that it is small when compared to fGn time series ($N = 500$). Climacograms were calculated on samples $\bar{X}_k^{(n)}$ ($n = N/10$), which do not cover the whole series when the ratio between series length N and the scale k is a non-integer. When this is the case, some of the time series values are not considered in the climacogram. In this way, the proposed alternative is to calculate the linearly sampled climacogram for the original one and the inverted inflow time series and calculating the mean between the two climacograms.

The study of the Hurst exponent can be helpful in the analysis of long-range behaviour for inflow time series, but it is necessary to consider that different methods can lead to different results, as shown in this work.

For the 148 HPPs, the climacogram based the Hurst exponent shows higher values, with a ratio of $H_{\text{clim}}/H_{R/S} = 1.2 \pm 0.1$. The following average Hurst exponents were determined: $\bar{H}_{R/S} = 0.58 \pm 0.06$ and $\bar{H}_{\text{clim}} = 0.70 \pm 0.10$. Nine HPP’s are non-persistent, with $H_{R/S} \lesssim 0.5$, and only four have $H_{\text{clim}} \lesssim 0.5$. The analysis by hydrographic regions has shown $H_{R/S}$ with a decreasing trend from upstream to downstream, a behaviour not seen for H_{clim} .

Taking arbitrary threshold ranges, $\bar{H}_{R/S} = 0.60$ and $\bar{H}_{\text{clim}} = 0.72$, it was verified that, for both methods, $H < \bar{H}$ for six hydrographic regions (Amazon, Eastern Atlantic, Paraná, Southeastern Atlantic, Southern Atlantic, and Uruguay) and that $H > \bar{H}$ for the four remaining regions (Tocantins-Araguaia, Paranaíba, São Francisco, and Paraguay).

The methods are applied to 50 river head annual inflow time series. The inflows show persistence in all cases for H_{clim} , and in all but three time series, for $H_{R/S}$. The behaviour of H as a function of localization depends on the method. The exception is the Uruguay hydrographic region, which shows the lowest values for both $H_{R/S}$ and H_{clim} . No pattern is detected when the analysis is done as a function of the type of HPP.

Bubble maps indicate the variation of the Hurst exponent over the Brazilian territory. No pattern is seen, except for a trend of smaller H values in the far north and far south regions.

The Pearson correlation coefficient between $H_{R/S}$ and H_{clim} indicates that the methods are similar, but do not present the same behaviour. For eight hydrographic regions, the coefficients are in the range $0.6 < r_p < 1$ (moderate to strong correlation), but two regions show weak (São Francisco) or no correlation at all (Uruguay), being more sensitive to the Hurst exponent method. This indicates that the inflow time series can be governed by local features.

The analysis of the Paraná hydrographic region shows that six HPPs in the Paranaíba basin (“Paranaíba subgroup”) and the four HPPs located on the Paraná River have the highest Hurst exponents ($H_{\text{clim}} \geq 0.8$). Only the HPPs at the Paranaíba basin showed $H_{R/S} > 0.6$.

Considering that the Hurst exponent is an important parameter for long-term behaviour analysis, this work shows that there is some degree of dependence on the method selected to perform it. The results show a global difference of 20% between the R/S analysis and the climacogram method. Nevertheless, it is seen that the inflow time series are persistent ($H > 0.5$) with few non-correlated exceptions.

The climacogram analysis can be improved with the linear sampling (avoiding overweighting of the larger scales) and using the inversion of time series order (which mitigates the effect of unanalysed points).

Acknowledgements

The authors would like to thank the coordination of the PPGERHA programme and the staff of the Hydraulics and Sanitation Department at Universidade Federal do Paraná.

This work was mostly developed with the support of the energy generation and transmission utility company in the State of Paraná, “Copel Geração e Transmissão”, under Project LYNX (ANEEL PD-06491-0307-2013). Many thanks to the project technical team and a special acknowledgement to the project manager, Klaus de Geus.

The authors would also like to thank Emerson L. Alberti, M.Eng. from “Centrais Elétricas do Rio Jordão” (Elejor) for the support under project ANEEL PD-02945-0001-2019.

References

- [1] H. E. Hurst. Long-Term Storage Capacity of Reservoirs. **Transactions of the American Society of Civil Engineers**, v. 116, pp. 770-808, 1951.
- [2] L. Seuront. **Fractals and Multifractals in Ecology and Aquatic Science**. CRC Press: Boca Raton, FL. 2010, (1st paperback ed.: 2017).
- [3] A. Eke, P. Herman, L. Kocsis, L. R. Kozak. Fractal characterization of complexity in temporal physiological signals. **Physiological Measurement**, v. 23(1), pp. R1-R38, 2002.
- [4] B. B. Mandelbrot & J. R. Wallis. Noah, Joseph, and Operational Hydrology. **Water Resources Research**, v. 4, pp. 909-918, 1968.
- [5] P. E. O’Connell, D. Koutsoyiannis, H. F. Lins, Y. Markonis, A. Montanari, T. Cohn. The scientific legacy of Harold Edwin Hurst (1880-1978). **Hydrological Sciences Journal**, v. 61(9), pp. 1571-1590, 2016.
- [6] B. B. Mandelbrot & J. W. Van Ness. Fractional Brownian Motions, Fractional Noises and Applications. **SIAM Review**, v. 10, n. 4, pp. 422-437, 1968.
- [7] D. Koutsoyiannis, A. Montanari. Statistical analysis of hydroclimatic time series: Uncertainty and insights. **Water Resources Research**, vol. 43, W05429, 2007.
- [8] T. Dieker. **Simulation of fractional Brownian motion**. Master’s thesis – revised. University of Twente. Online document accessed: Feb/01/2021. Available at: <http://www.columbia.edu/~ad3217/fbm/thesisold.pdf>. 2004.
- [9] MathWorks®. Online documentation available at: <https://www.mathworks.com/help/wavelet/ref/wfbm.html>.
- [10] P. Abry, F. Sellan. The Wavelet-Based Synthesis for Fractional Brownian Motion Proposed by F. Sellan and Y. Meyer: Remarks and Fast Implementation. **Applied and Computational Harmonic Analysis**, v. 3(4), pp. 377-383, 1996.
- [11] J.-M. Bardet, G. Lang, G. Oppenheim, A. Philippe, S. Stoev, M. S. Taqqu. **Generators of Long-Range Dependent Processes: A Survey**. In: Theory and Applications of Long-Range Dependence, P. Doukhan, G. Oppenheim, M. S. Taqqu (eds.), pp. 579-623. Boston: Birkhauser, 2003.
- [12] OriginLab Technical Support. App for Hurst exponent determination with R/S analysis. Available at: www.originlab.com/FileExchange/details.aspx?fid=289.
- [13] R. Weron. Estimating long-range dependence: finite sample properties and confidence intervals. **Physica A**, v. 312, pp. 285-299, 2002, doi: 10.1016/S0378-4371(02)00961-5.
- [14] R. Weron. **HURST: MATLAB function to compute the Hurst exponent using R/S Analysis**. HSC Software M11003, Hugo Steinhaus Center, Wroclaw University of Technology, 2011. Online documentation available at: <https://ideas.repec.org/c/wuu/hscode/m11003.html>.
- [15] D. Koutsoyiannis. Nonstationarity versus scaling in hydrology. **Journal of Hydrology**, v. 324, 239-254, 2006.
- [16] G. Poulialis. Digital documentation available at <https://github.com/Gapoulialis/Stochastic-Models/blob/master/Climacogram.m>.
- [17] P. Dimitriadis, D. Koutsoyiannis. Climacogram versus autocovariance and power spectrum in stochastic modelling for Markovian and Hurst-Kolmogorov processes. **Stochastic Environmental Research and Risk Assessment**, v. 29, pp. 1649-1669, 2015.
- [18] P. Dimitriadis, D. Koutsoyiannis. The mode of the climacogram estimator for a Gaussian Hurst-Kolmogorov process. **Journal of Hydroinformatics**, v. 22(1), pp. 160-169, 2020.
- [19] Nile River time series are available in several databases. The Project R user can access the data at https://rstudio-pubs-static.s3.amazonaws.com/574608_9b4f5b489c3a4ec8a09163b1b470db12.html.
- [20] D. Koutsoyiannis. Climacogram-based pseudospectrum: a simple tool to assess scaling properties. **European Geosciences Union General Assembly 2013**. Vienna, Austria, April 2013.
- [21] H. Wen, Z. Liu. Separating Fractal and Oscillatory Components in the Power Spectrum of Neurophysiological Signal. **Brain Topography**, v. 29, pp. 13-26, 2016.
- [22] H. Wen, Z. Liu. **Separating Fractal and Oscillatory Components in the Power Spectrum of Neurophysiological Signal**. Purdue University Research Repository. doi: 10.4231/R7WQ01R7. Version 1.0, 2015.
- [23] ONS. **Procedimentos de Rede (Network Procedures). Submodule 2.4: Criteria for Energetic and Hydrological Studies**. Online documentation available at: <http://www.ons.org.br/paginas/sobre-o-ons/procedimentos-de-rede/vigentes>.