ISSN: 2447-3359

REVISTA DE GEOCIÊNCIAS DO NORDESTE

Northeast Geosciences Journal

v. 11, nº 1 (2025)

https://doi.org/10.21680/2447-3359.2025v11n1ID37090

Statistical properties of historical water level and flow rate series of the doce river basin

Propriedades estatísticas das séries históricas de cota e vazão da bacia do rio Doce

Emmanuel Kennedy Costa Teixeira [1]

[1] Federal University of Viçosa, Department of Civil Engineering, Viçosa/MG, Brazil. Email: emmanuel.teixeira@ufv.br
ORCID: https://orcid.org/0000-0001-7598-0240

Abstract: The Doce River Basin holds immense socio-environmental, economic, and cultural significance, playing a vital role in the lives of local communities, regional biodiversity, and the preservation of water resources. For water resource planning, historical data series are commonly used. However, these series are not always statistically analyzed prior to use. Therefore, this study employed non-parametric tests using the statistical software R to assess the randomness, independence, homogeneity, and stationarity of the monthly mean water level and flow rate series of this basin. The results indicated that none of the stations had random series; four stations had independent series (three for water level and one for flow rate); 128 stations had homogeneous series (58 for water level and 70 for flow rate); and 148 stations had stationary series (76 for water level and 72 for flow rate). In conclusion, the historical data series exhibit consistent variations and should be thoroughly analyzed before being utilized.

Keywords: Non-parametric tests; Hydrometric stations; Statistical hydrology.

Resumo: A bacia do rio Doce é de extrema importância socioambiental, econômica e cultural, desempenhando um papel vital na vida das comunidades locais, na biodiversidade regional e na preservação dos recursos hídricos. Para realizar o planejamento dos seus recursos hídricos, normalmente se recorre às séries históricas de dados. Entretanto, nem sempre elas são analisadas estatisticamente antes de serem utilizadas. Sendo assim, este estudo verificou, por meio da aplicação de testes não paramétricos, utilizando o programa estatístico R, a aleatoriedade, independência, homogeneidade e estacionariedade das séries de médias mensais de cota do nível d'água e de vazão desta bacia. A partir dos resultados, obteve-se que nenhuma estação possuía série aleatória; quatro possuíam séries independentes (três de cota e uma de vazão); cento e vinte e oito homogêneas (58 de cota e 70 de vazão); cento e quarenta e oito estacionárias (76 de cota e 72 de vazão). Ao final, concluiu-se que os dados das séries históricas possuem constâncias em suas variações e devem ser analisados criteriosamente antes de serem utilizados.

Palavras-chave: Testes não paramétricos; Estações fluviométricas; Hidrologia estatística.

Recebido: 26/07/2024; Aceito: 10/01/2025; Publicado: 05/02/2025.

1. Introduction

Water resources are crucial for the preservation of the life cycle. Therefore, one of the major challenges today is their sustainable management. Badham et al. (2019) suggest that planning should be done in a way that the interaction between humans and the environment does not harm the ecosystem. It is necessary that projects use data aligned with the reality of the hydrological region, ensuring effective management and preservation of these resources.

In this context, Silva (2020) emphasizes that statistical analysis of historical flow data is essential for water use planning, hydropower development, flood forecasting, watershed management, among others. Uliana et al. (2015) stated that this analysis can be done through non-parametric tests, which, according to Naghettini (2017), can assess the randomness, independence, homogeneity, and stationarity of hydrological series data.

Several studies have analyzed historical flow data in watersheds, including Câmara et al. (2016), who tested and identified the non-stationarity of flow data from the Tocantins River. This highlights the need to consider these changes in hydrological models.

Carvalho and Ruiz (2016) conducted a homogeneity study to analyze and fill data gaps using the Simple or Multiple Linear Regression Method for 19 rainfall stations in the Araguari River basin. Through statistical tests, such as the Pettitt test and the Mann-Kendall test, they identified significant changes in the data of some series, which could be related to factors like changes in measurement instruments, station location shifts, or regional climate variations.

One of the most important watersheds in the Southeast region is the Doce River basin. It holds significant socioeconomic and environmental relevance for local communities. Additionally, it provides water resources for supply, agriculture, industry, and other economic sectors, while also being crucial for the preservation of biodiversity and regional ecosystems. Therefore, for the protection of its water resources, planning is essential, for which historical data series are used.

Historically, the basin has suffered from flooding and inundation events, which have impacted the cities located along its course. Almeida (2020) stated that floods have disrupted economic activities established along the basin, notably agriculture, mining, industry, and energy generation, as reported by Barbosa (2004). This underscores the need to use historical series for planning in the basin to address flood periods.

Several authors have conducted statistical analyses in the Doce River basin, such as Valverde et al. (2003), who adjusted statistical models to calculate the homogeneity of historical series and established equations to determine which regions of the Doce River were homogeneous or heterogeneous during periods of heavy rainfall.

Teixeira et al. (2018), aiming to determine the Manning roughness coefficients for various stretches along the Doce River, analyzed the randomness of flow data from flow stations in the basin. The historical series exhibited statistically random behavior, with no significant trends or seasonal patterns that could compromise the analyses.

Silva, Lima, and Groppo (2017) used the Mann-Kendall and Pettitt tests to analyze the independence and randomness of flow data from the Doce River, in order to determine temporal trends and identify possible abrupt changes in the river's average flow between 1950 and 2016. The results indicated that all stations exhibited negative trends in the flow time series.

Ferreira et al. (2020), in their analysis of the Guanhães River flows, tested stationarity to quantify and investigate the behavior and trends of the flows over the years, checking for possible behavioral changes. Fraga et al. (2020) applied the Pettitt test and linear regression analysis to evaluate the annual and monthly precipitation and flow data at stations in the Piracicaba River Water Resources Management Unit. The precipitation series were stable over time, with few significant trends, while the flow data showed reduction trends at several monitored stations.

Although there are studies on the Doce River basin, no materials have been found that statistically and extensively analyzed the historical series of water level and flow data. Heidarpour et al. (2017) stated that hydrological variables, such as flow and water levels, should be analyzed with precision to avoid excessive project costs, large-scale damage, or even loss of human lives. Additionally, Hodgkins et al. (2017) emphasized that data analysis is essential, as methods developed for estimating flow and water levels rely on statistical assumptions about the historical series used. Therefore, this study aimed to analyze the historical series from all flow stations in the Doce River basin, in order to test their randomness, independence, homogeneity, and stationarity. This contributes to better water resources planning in the basin and helps ensure that future studies do not adopt incorrect statistical assumptions.

2. Methodology

The methodology encompasses the following steps: (a) characterization of the study area; (b) acquisition of historical series data; and (c) statistical analysis of the historical series.

2.1 Study area

The Doce River basin is located in the Southeast region of Brazil, with a drainage area of approximately 86,715 km², of which 86% is located in the eastern part of the state of Minas Gerais and 14% in the northeastern part of Espírito Santo. Its headwaters are in the Espinhaço and Mantiqueira mountain ranges, located in the state of Minas Gerais. The basin predominantly experiences a tropical highland climate, with annual precipitation volumes ranging from 900 mm to 1,500 mm, and a landscape characterized by rolling, mountainous, and rugged terrain (ALMEIDA, 2020).

Flooding, which poses a significant problem for the basin, is recorded during the rainy season, occurring from October to March, with higher volumes in the months of December to February. Additionally, the unplanned occupation of the floodplain of watercourses, especially in urban areas, has worsened the damage caused by floods. To mitigate these damages, a flood warning system has been in operation in the basin for 13 years, since the 1997/1998 rainy season, highlighting the need and feasibility that the calculation and statistical analysis of historical series provide for understanding, error correction, and predicting the future behavior of the basin.

2.2 Acquisition of historical series data

Bayer, Castro, and Bayer (2012) stated that a time series is a set of observations ordered over time, making it possible to conclude that processes within the hydrological cycle, such as flow, among others, can be treated as this type of series, since they are systematically measured at discrete time intervals.

According to Naghettini and Pinto (2007), a hydrological series can include all available observations collected at regular intervals over several years or only some of its characteristic values, such as annual maxima or monthly averages. The statistical analysis of these data allows for identifying trends in occurrences.

The historical series analyzed in this study were obtained from the Hydrological Information System (HidroWeb) of the National Water and Basic Sanitation Agency (ANA). For the Doce River basin, data from 211 flow stations were collected, which, theoretically, had daily historical series of flow and water level records.

For the statistical analysis of the data, the main tool used was the R statistical programming software, which is freely accessible. The historical series values were imported into R. However, the spreadsheet obtained from HidroWeb contained the value “0” (zero) in the data gaps of the historical series. These gaps would be interpreted by R as zeroed data, which would affect the analysis. Therefore, before exporting the data to the statistical program, all zeros were replaced by “NA” (not available) in the spreadsheet, which R interprets as missing data.

Another important step was replacing the decimal separator, as HidroWeb data uses commas, while R recognizes only the period as a decimal separator. Additionally, before export, the data were verticalized, meaning a single column was created containing all the values, with each row having the daily data of the historical series from each flow station. Subsequently, the data were imported into the statistical program.

2.3 Statistical analysis of historical series

The statistical analyses of the daily historical series of flow and water level data from the flow stations in the Doce River basin were carried out using non-parametric tests. These tests were implemented in the R program, utilizing the equations for each test as described by Naghettini and Pinto (2007) and presented below.

For the randomness test, the hydrological data series was analyzed for variation due to natural causes. The test used in this study was the Inflection Number Test (Equation 1).

$$E(p) = \frac{2(N-2)}{3} \tag{1}$$

In which:

E(p) is the expected number of inflections;

N is the sample size (number of observations).

Considering that the null hypothesis (H0) is that the sample is random, the test statistic is represented by Equation 2.

$$T = \frac{p - E(p)}{\sqrt{Var(p)}} \quad (2)$$

In which:

T is the statistic of the non-parametric test;

p is the number of inflections observed in the data;

Var (p) is the variance of the number of inflections (Equation 3).

$$Var(p) = \frac{16N - 29}{90} \quad (3)$$

With a significance level of 5%, the null hypothesis is rejected if $|T| > Z_{97,5}$ (two-tailed test).

The independence test examined the possibility that the data in the sample could influence any previous observations. To perform this verification, the Wald and Wolfowitz test was used. Considering that the null hypothesis assumes the sample is independent, the test statistic is represented by Equation 4.

$$T = \frac{R - E[R]}{\sqrt{VAR[R]}} \quad (4)$$

In which:

T it is the test statistic for the non-parametric test;

R is a coefficient of the Wald and Wolfowitz test;

$E[R]$ it is the mean of the coefficient R;

$Var[R]$ it is the variance of the coefficient R.

The coefficient R was calculated using Equation 5.

$$R = \sum_{i=1}^{N-1} X_i' X_{i+1}' + X_1' X_N' \quad (5)$$

In which:

N is the sample size;

X_i' is the difference between a data point in the sample at position i and the sample mean of the data.

If the historical series is independent, the coefficient R follows a normal distribution with a mean *E[R]* calculated according to Equation 6.

$$E[R] = -\frac{s_2}{N - 1} \quad (6)$$

Where s is a variable of the test, calculated using Equations 7 and 8.

$$s_r = N m_r' \quad (7)$$

$$m_r' = \frac{\sum_{i=1}^{N} (X_i')^r}{N} \quad (8)$$

r denotes the order of the sample moments relative to the origin. The variance of the coefficient R is given by Equation 9.

$$Var[R] = \frac{s_2^2 - s_4}{N-1} + \frac{s_2^2 - 2s_4}{(N-1)(N-2)} - \frac{s_2^2}{(N-1)^2} \quad (9)$$

With a significance level of 5%, the null hypothesis is rejected if $|T| > Z_{97,5}$ (two-tailed test).

To test the homogeneity of the historical series, the data were examined to determine if they were similar or originated from the same population. For this, the non-parametric test proposed by Mann and Whitney was applied. Considering the data from any sample of size N $\{X_1, X_2, X_3, ..., X_n\}$, which was divided into two sub-samples, where $\{X_1, X_2, X_3, ... X_{N1}\}$ of size N_1 e $\{X_{N+1}, X_{N+2}, X_{N+3}, ... X_N\}$ of size N_2, such that $N_1 + N_2 = N$. The sizes of the sub-samples should be approximately equal, such that $N_1 \leq N_2$. The complete sample (size N) was ordered in ascending order, and its elements were assigned an ordinal rank m.

The coefficient V of the test is given by the smallest value between Equations 10 and 11.

$$V_1 = N_1N_2 + \frac{N_1(N_1+1)}{2} - R_1 \quad (10)$$

$$V_2 = N_1N_2 - V_1 \quad (11)$$

In which:

V_1 e V_2 they are coefficients of the Mann-Whitney test.;

N_1 it is the size of sub-sample 1;

N_2 it is the size of sub-sample 2;

R_1 it is the sum of the rankings of the data in the first sub-sample (size $N1$).

If the historical series is homogeneous and $N_1, N_2 > 20$, the coeficiente V follows a normal distribution with a mean calculated using Equation 12.

$$E[V] = \frac{N_1N_2}{2} \quad (12)$$

$E[V]$ it is the mean of the coefficient V.

The variance of the coefficient V ($Var[V]$) was calculated using Equation 13.

$$VAR[V] = \frac{N_1N_2(N_1+N_2+1)}{12} \quad (13)$$

The test statistic T is represented by Equation 14.

$$T = \frac{V - E[V]}{\sqrt{VAR[V]}} \quad (14)$$

The null hypothesis, which assumes that the sample is homogeneous, is rejected at a 5% significance level if $|T| > Z_{97,5}$ (two-tailed test).

Regarding stationarity, Bauzha and Gorbachova (2017) stated that the test is based on the assumption that the flow generation process is in equilibrium around an underlying mean and that its variation remains constant over time. The stationarity test examined whether the data, excluding random variations, did not exhibit changes. In this study, the Spearman test was used. Considering the data of any sample of size N $\{X_1, X_2, X_3, ..., X_n\}$, the correlation coefficient between the occurrence order T_t and the ascending rank order m_t of the elements was analyzed.

The Spearman correlation coefficient r_s was calculated using Equation 15.

$$r_s = 1 - \frac{6\sum_{t=1}^{N}(m_t - T_t)^2}{N^3 - N} \quad (15)$$

In which:

T_t, it is the order of occurrence of the data;

m_t it is the increasing ranking of the data.

If the historical series has a size $N > 10$, the coefficient r_s follows a normal distribution, with its mean ($E[r_s]$) obtained from Equation 16 and its variance ($VAR[r_s]$) from Equation 17.

$$E[r_s] = 0 \tag{16}$$

$$VAR[r_s] = \frac{1}{N-1} \tag{17}$$

Considering that the null hypothesis is that T_t and m_t are not correlated, the test statistic T is represented by Equation 18.

$$T = \frac{r_s}{\sqrt{VAR[r_s]}} \tag{18}$$

If $|T| > Z_{97,5}$ (two-tailed test), the null hypothesis is rejected at a 5% significance level. After the historical series data were imported into R, a function was implemented to perform all the non-parametric tests mentioned. This function received as input parameters the vector with the data, the start year, and the end year of each series.

The data obtained from HidroWeb are daily; however, the tests were applied to monthly averages. This was done because average flow rates are the most important for basin planning, as they represent water availability. They are also used for determining the energy potential of the basin and in flow regulation.

Thus, the first part of the implemented function aimed to return a matrix with the monthly average flow rates for each hydrometric station. Separately, the function developed in R also returned a matrix with the monthly average water levels for each station. The function performed a day-by-day, month-by-month, and year-by-year count. Therefore, it was necessary for the first data imported into R to be January 1st of the start year of the series. Leap years were also taken into account.

The tests were not applied to the hydrometric stations that presented few recorded data, as unreliable results would be obtained. Morettin and Bussab (2010) stated that, for the test results to be reliable, the sample size must be greater than 30, which was the value adopted in this work.

As mentioned, the implementation of the tests (randomness, independence, homogeneity, and stationarity, in this order) was carried out through a function developed in R. Thus, the monthly data from the generated matrix were subsequently tested, and for each test applied, the function displayed the following message on the screen: *"The null hypothesis (H_0) is rejected at a 5% significance level," or "The null hypothesis (H_0) is not rejected at a 5% significance level."* Additionally, if the number of data points was below the minimum required for the test, the function would display the warning: *"WARNING! The available data is insufficient for a reliable test."*

Since there were many historical series to be tested, as 211 gauging stations in the Doce River basin had flow and water level data, a function was developed in R to automatically read the hundreds of files and subsequently apply the non-parametric tests. At the end, the function also automatically recorded the obtained results.

3. Results and discussion

After organizing the data and applying the function developed in the R program, it was found that, of the 211 stations in the Doce River basin, four (1.90%) did not have water level gauge data available, and 12 (5.69%) had insufficient data for the application of the tests. For the remaining 195 stations, the average evaluation period was 26.91 years, with the longest period being 91 years.

Regarding the analysis of average flow rates, 70 stations had unavailable data, which corresponds to 33.18%, and five (2.37%) had insufficient data for the application of hypothesis tests. On average, the stations had 20.42 years of recorded data, with the longest period being 86 years.

Thus, as it was observed that many stations either did not have recorded data in their historical series or had insufficient data for the application of non-parametric tests, it is emphasized that it is important and necessary for future studies in the basin to address the data gaps. Carvalho and Ruiz (2016) reported that filling these gaps assists in the development of projects that rely on the analysis of extensive historical series.

The non-parametric tests were conducted individually for each gauging station, for the monthly flow and water level data. In Table 1, the column "S" indicates the series that exhibited randomness, independence, homogeneity, and/or stationarity, while "N" indicates that the data did not follow these characteristics.

Table 1 – Results of the non-parametric tests applied to the historical series of monthly flow and water level data from the gauging stations in the Doce River basin

Statistical property	Variable	S	N
Randomness	Water level	0	195
	Flow rate	0	136
Independence	Water level	3	192
	Flow rate	1	135
Homogeneity	Water level	58	137
	Flow rate	70	66
Stationarity	Water level	76	119
	Flow rate	72	64

Source: Author (2023).

Based on the results from Table 1, it is observed that no gauging station exhibited a random series for the monthly mean water level and discharge data. This indicates that the hydrological data do not show peaks of variation due to natural causes. Ferreira et al. (2020), for a gauging station in the Doce River basin, also found that the variation in discharge is anthropogenic, linked to land use and occupation.

Regarding independence, a small number of gauging stations exhibited this characteristic in their monthly mean water level (three stations) and discharge (one station) series. The stations with independent water level data were codes 56846890, 56415000, and 56610000. The station with independent discharge data was code 56661000.

As for stationarity, a greater number of gauging stations displayed this characteristic, with 76 stations (36.02% of the total in the Doce River basin) having stationary monthly mean water level data and 72 stations (34.12%) for discharge. The stationary stations are listed in Table 2.

Table 2 – Stations with stationary series.

Average water level quotas
56695000, 56991000, 56681000, 56989500, 56987900, 56990500, 56993550, 56992500, 56719998, 56993300, 56565000, 56055000, 56630000, 56230000, 56974000, 56997200, 56780000, 56540001, 56635000, 56470000, 56925001, 56640000, 56182000, 56994500, 56994505, 56994502, 56943000, 56935000, 56340000, 56846890, 56590000, 56976000, 56847000, 56846900, 56246000, 56999000, 56380000, 56090000, 56185000, 56744500, 56510000, 56993005, 56993002, 56992900, 56400000, 56998100, 56998005, 56989001, 56989000, 56350000, 56659999, 56207000, 56963000, 56835000, 56633000, 56195000, 56105000, 56995500, 56995500, 56430000, 56631500, 56820000, 56350000, 56210000, 56948005, 56978000, 56924100, 56924000, 56919500, 56927000, 56015000, 56650000, 56170000, 56520000, 56892000, 56996000
Average flow rates
56500000, 56695000, 56991000, 56990990, 56681000, 56992000, 56997000, 56990500, 56940000, 56719998, 56336001, 56565000, 56974000, 56012000, 56540001, 56539000, 56720000, 56635000, 56470000, 56640000, 56985000, 56182000, 56994502, 56943000, 56700000, 56935000, 56983000, 56590000, 56976000, 56847000, 56846900, 56005000, 56780005, 56670000, 56880000, 56960005, 56185000, 56850000, 56988500, 56993002, 56400000, 56991500, 56155000, 56998100, 56960000, 56150000, 56696000, 56460000, 56989001, 56989000, 56825000, 56659999, 56659998, 56661000, 56152000, 56963000, 56982000, 56570000, 56633000, 56195000, 56105000, 56995500, 56995500, 56430000, 56631500, 56160000, 56158000, 56485000, 56948005, 56385000, 56990000, 56065000, 56130000

Source: Author (2023).

Although the number of stations exhibiting this characteristic is higher compared to others, stationarity should be analyzed with greater statistical rigor, as recommended by Li et al. (2022). This is because several factors, such as climate change and human activities, can influence the trend of the data. Therefore, it is recommended that more than one statistical test be applied to confirm the hypothesis of stationarity.

According to Table 2, it was observed that 27.49% of the stations had homogeneous monthly mean series for water level quotas and 33.18% for flow rates. These series exhibited small variations in flow rates and water levels over the examined period. This indicates the basin's capacity to provide water resources and support, as well as the feasibility of urban occupation and expansion in its surroundings. The codes of the stations with homogeneous monthly mean series for water level quotas and flow rates are presented in Table 3.

Table 3 – Stations with homogeneous series.

Average water level quotas
56695000, 56991000, 56990990, 56050000, 56760000, 56989500, 56989400, 56993550, 56992500, 56719998, 56993300, 56565000, 56055000, 56630000, 56230000, 56997200, 56995000, 56780000, 56635000, 56470000, 56182000, 56994500, 56994510, 56994502, 56750000, 56935000, 56340000, 56590000, 56986000, 56976000, 56846900, 56380000, 56240000, 56775000,56744500, 56510000, 56993005, 56993002, 56400000, 56998100, 56998000, 56989001, 56350000, 56028000, 56633000, 56195000, 56105000, 56995500, 56995500, 56631500, 56947000, 56948005, 56415000, 56015000, 56650000, 56170000, 56520000, 56892000
Average flow rates
56991000, 56990990, 56989400, 56990500, 56940000, 56719998, 56565000, 56055000, 56012000, 56995000, 56539000, 56720000, 56900000, 56635000, 56470000, 56985000, 56182000, 56994510, 56994502, 56632000, 56750000, 56935000, 56340000, 56983000, 56590000, 56986000, 56976000, 56846900, 56005000, 56337000, 56880000, 56960005, 56090000, 56510000, 56988500, 56400000, 56991500, 56155000, 56998100, 56960000, 56150000, 56696000, 56989001, 56659999, 56659998, 56963000, 56982000, 56570000, 56028000, 56633000, 56195000, 56995500, 56995500, 56430000, 56631500, 56160000, 56110005, 56158000, 56485000, 56948005, 56415000, 56978000, 56385000, 56990000, 56065000, 56085000, 56130000, 56170000, 56520000, 56892000

Source: Author (2023).

Evaluating the stations that presented positive results for the analyses, it was found that 35 were stationary both for flow and water level, corresponding to 16.59% of the total number of hydrometric stations in the Doce River Basin. Additionally, 34 stations (16.11%) had historical series that were homogeneous for both flow and water level. In the case of randomness and independence, no station showed a positive result simultaneously for both water level and flow.

It is observed that, although water level and flow are directly related, there were cases where a single hydrometric station had a stationary series for flow but not for water level, and vice versa. The same was observed for homogeneity and independence. This reveals the deficiencies and uncertainties inherent in flow and water level monitoring techniques. Moreover, it is important to note that each hypothesis test has its power, meaning there is a probability of rejecting the null hypothesis when it is true.

Based on the analyses, it was observed that, for both water level and flow, no station presented data that were simultaneously random, independent, homogeneous, and stationary. Therefore, assuming these premises, which are typical in hydrological studies, may be misleading for most series, which, in turn, produces unrepresentative statistics and affects the quality and outcome of the project. Thus, studies should apply a statistical test to verify the hypothesis, rather than simply assuming it to be true. If the premise is not accepted, techniques can be applied to correct trends for current conditions.

4. Final considerations

Throughout this study, the foundations guiding the analysis of randomness, independence, stationarity, and homogeneity of monthly averages of discharge rates and water level elevations from the historical series of gauging stations in the Doce River Basin were presented. Based on the results obtained, the following conclusions can be drawn:

- Of the 211 stations examined, approximately 7.5% did not have monthly average data for water level elevations, or the data were insufficient for the application of hypothesis tests. In the case of discharge rates, this figure increased to approximately 35.5% of the gauging stations.

- No station presented random monthly average data, either for water level elevations or discharge rates. Regarding independence, three stations exhibited independence for water level elevations, and only one station for discharge rates.

This indicates that the data from the historical series exhibit constant variations without following a specific pattern. Furthermore, there is evidence that the hydrological behavior has been influenced by anthropogenic activities, such as the construction of dams.

- For homogeneity, 58 and 70 gauging stations exhibited this characteristic for water level elevation and discharge data, respectively.
- A total of 76 stations had stationary data for water level elevation, and 72 for discharge.
- Although water level elevation and discharge are directly related variables, only 35 stations were stationary for both parameters simultaneously. In the case of homogeneity, this number decreased to 34. Therefore, it can be observed that the variations in the data may be related to uncertainties and errors in their measurement.
- No station presented a series that was simultaneously random, independent, stationary, and homogeneous, both for average discharge and average water level elevation. Thus, hydrological studies to be conducted in the basin should take this into account before assuming these four premises.

For future work, it is suggested that the methodology presented in this paper be applied to verify randomness, independence, homogeneity, and stationarity in the series of minimum and maximum discharge rates in the basin. Additionally, it would be valuable to evaluate other river basins.

References

Almeida, F. C. Análise multicritério na definição de áreas prioritárias para recuperação florestal na bacia do Rio Doce, em Minas Gerais. *Nativa*, v. 8, n. 1, 81-90, 2020. http://dx.doi.org/10.31413/nativa.v8i1.8130.

Badham, J.; Elsawah, S.; Guillaume J. H. A.; Hamilton, S. H.; Hunt, R. J.; Jakeman, A. J.; Pierce, S. A.; Snow, V. O.; Babbar-Sebens, M.; Fu, B.; Gober, P.; Hill, M. C.; Iwanaga, T.; Loucks, D. P.; Merritt W. S.; Peckham, S. D.; Richmond, A. K.; Zare, F.; Ames, D.; Bammer, G. Effective modeling for Integrated Water Resource Management: a guide to contextual practices by phases and steps and future opportunities. *Environmental Modelling and Software*, v. 116, 40–56, 2019. https://doi.org/10.1016/j.envsoft.2019.02.013.

BARBOSA, S. E. S. *Análise de dados hidrológicos e regionalização de vazões da bacia do rio do Carmo*. Ouro Preto, 2004.188f. Dissertação (Mestrado em Engenharia Ambiental). Programa de Pós-Graduação em Engenharia Ambiental da Universidade Federal de Ouro Preto, Ouro Preto, 2004.

Bauzha, T.; Gorbachova, L. The features of the cyclical fluctuations, homogeneity and stationarity of the average annual flow of the southern Buh River basin. *Annals of Valahia University of Targoviste. Geographical Series*, v. 17, n. 1, 5-17, 2017. https://doi.org/10.1515/avutgs-2017-0001.

Bayer, D. M.; Castro, N. M. R.; Bayer, F. M. Modelagem e previsão de vazões médias mensais do rio Potiribu utilizando modelos de séries temporais. *Revista Brasileira de Recursos Hídricos*, v. 17, n. 2, 229-239, 2012. https://doi.org/10.21168/rbrh.v17n2.p229-239.

Câmara, R. K. C.; Rocha, E. J. P.; Protázio, J. M. B.; Queiroz, J. C.; Ribeiro, W, M, N.; Siqueira, I. S.; Lima, A. M. M. Modelagem hidrológica estocástica aplicada ao rio Tocantins para a cidade de Marabá - PA. *Revista Brasileira de Meteorologia*, v. 31, n. 1, 11-23, 2016. https://doi.org/10.1590/0102-778620140092.

Carvalho, H. P.; Ruiz, M. V. S. Avaliação da consistência de séries históricas de chuva da bacia hidrográfica do Rio Araguari, em Minas Gerais**.** *Periódico Eletrônico Fórum Ambiental da Alto Paulista*, v. 12, n. 6, 2016.

COSTA, K. T. *Avaliação de distribuições de probabilidades das vazões médias diárias máximas anuais do Brasil.* Belo Horizonte, 2014. 212f. Dissertação (Mestrado em Saneamento, Meio Ambiente e Recursos Hídricos). Programa de Pós-Graduação em Saneamento, Meio Ambiente e Recursos Hídricos, Universidade Federal de Minas Gerais, Belo Horizonte, 2014.

Detzel, D. H. M.; Bessa, M. R.; Vallejos, C. A. V.; Santos, A. B.; Thomsen, L. S.; Mine, M. R. M.; Bloot, M. L.; Estrócio, J. P. Estacionariedade das afluências às usinas hidrelétricas brasileiras. *Revista Brasileira de Recursos Hídricos*, v. 16, n. 3, 95-111, 2011. https://doi.org/10.21168/rbrh.v16n3.p95-111.

Fraga, M. S.; Abreu, M. C.; Reis, G. B.; Tozi, T.; Pinheiro, S. A. R. Análise de tendência em séries históricas de precipitação e vazão na UGRH2 Piracicaba, Minas Gerais. *Revista Ibero-Americana de Ciências Ambientais*, v. 11, n. 2, 1-9, 2020. https://doi.org/10.6008/CBPC2179-6858.2020.002.0016.

Ferreira, F. L. V.; Rodrigues, L. N.; Almeida, L. T.; Teixeira, D. B. S. Tendência em séries hidrológicas e de mudanças no uso e cobertura da terra da bacia hidrográfica do rio Guanhães, Minas Gerais. *Brazilian Journal of Animal and Environmental Research*, v. 3, n. 2, 447-459, 2020. https://doi.org/10.34188/bjaerv3n2-004.

Heidarpour, B.; Saghafian, B.; Yazdi, J.; Azamathulla, H. M. Effect of extraordinary large floods on at-site flood frequency. *Water Resources Management*, v. 31, 4187-4205, 2017. https://doi.org/10.1007/s11269-017-1739-x.

Hodgkins, G. A.; Whitfield, P. H.; Burn, D. H. Climate-driven variability in the occurrence of major floods across North America and Eutorpe. *Journal of Hydrology*, v. 552, 704-717, 2017. https://doi.org/10.1016/j.jhydrol.2017.07.027.

Li, Q.; Zeng, H.; Liu, P.; Li, Z.; Yu, W.; Zhou, H. Bivariate nonstationary extreme flood risk estimation using mixture distribution and copula function for the longmen reservoir, north China. *Water,* v.14, n. 4, 1-16, 2022. https://doi.org/10.3390/w14040604.

Morettin, P. A.; Bussab, W. O. *Estatística básica*. São Paulo: Saraiva, 2010. 6 ed.

Naghettini, M. *Fundamentals of statistical hydrology*. Switzerland: Springer International Publishing, 2017.

Naghettini, M.; Pinto, E. J. A. *Hidrologia estatística*. Belo Horizonte: CPRM, 2007.

Silva, L. E. F. Análise de tendência em séries históricas de vazão: uso de teste estatístico paramétrico. *Brazilian Applied Science Review*, v. 4, n. 3, 998-1018, 2020. https://doi.org/10.34115/basrv4n3-020.

Silva, R. K. F.; Lima, E. K. N.; Groppo, J. D. Análise de tendência das séries temporais de vazão na bacia do Rio Doce. In: CONGRESSO NACIONAL DE MEIO AMBIENTE - POÇOS DE ÁGUAS TERMAIS E MINERAIS, 2017, Poços de Caldas-MG. **Anais**...IFSULDEMINAS, 2017. p. 1-3. Disponível em: http://www.meioambientepocos.com.br/anais-simposio/anais-simposio/trabalhos/273.pdf. Acesso em: 23 de maio de 2022.

Teixeira, E. K. C.; Coelho, M. M. L. P.; Pinto, E. J. A.; Diniz, J. G.; Saliba. A. P. M. Coeficiente de rugosidade de Manning para o rio Doce. *Revista Brasileira de Recursos Hídricos*, v. 23, 1-12, 2018. https://doi.org/10.1590/2318-0331.231820180013.

Uliana, E. M.; Silva, D. D.; Uliana, E. M.; Rodrigues, B. S.; Corrêdo, L. P. Análise de tendência em séries históricas de vazão e precipitação: uso de teste estatístico não paramétrico. *Revista Ambiente e Água*, v. 10, n. 1, 1-7, 2015. https://doi.org/10.4136/ambi-agua.1427.

Valverde, A. E. L.; Silva, D. D.; Pruski, F. F.; Leite, H. G; Brandão, V. S. Análise regional de chuvas intensas para a bacia do Rio Doce. *Revista Brasileira de Recursos Hídricos*, v. 8, n. 4, 157-168, 2003. https://doi.org/10.21168/rbrh.v8n4.p157-168.