Revista Brasileira de Geografia Física

Homepage: <https://periodicos.ufpe.br/revistas/rbgfe>

Low flows hydrological regionalization and definition of homogeneous regions using multivariate statistical analyses in the Uruguai River watershed, on the Rio Grande do Sul State side, Brazil

Mylena Feitosa Tormam¹, Hugo Alexandre Soares Guedes², Carina Krüger Bork³, Micael de Souza Fraga⁴

¹ Engenheira Civil, Gerência de Recursos Hídricos e Meteorologia, Operador Nacional do Sistema Elétrico, Rua Júlio do Carmo, 251, Centro, CEP 20211-160, Rio de Janeiro, Rio de Janeiro. (53) 98468-8006. mylenatormam@ons.com.br. ² Professor Dr. Associado I, Universidade Federal de Pelotas, Centro de Engenharias, Rua Benjamin Constant, 989, Centro, CEP 96010-020, Pelotas, Rio Grande do Sul. (53) 3284-1614. hugo.guedes@ufpel.edu.br (autor correspondente). ³ Msc. em Recursos Hídricos, Programa de Pós Graduação em Recursos Hídricos, Unidade Federal de Pelotas, Rua Gomes Carneiro, 01, Andar Térreo, Bloco B, Sala 122, Centro, CEP 96010-610, Pelotas, Rio Grande do Sul. (53) 3284-3841. borkcarina@gmail.com. ⁴ Dr. em Engenharia Agrícola, Gerência de Monitoramento Hidrometeorológico e Eventos Críticos, Instituto Mineiro de Gestão das Águas, Rodovia João Paulo II, 4143 CEP 31630-900, Belo Horizonte, Minas Gerais. (31) 3915-1254. micaelfraga@gmail.com.

Artigo recebido em 09/09/2020 e aceito em 11/12/2020

ABSTRACT

Low flow estimation is essential for water resource management and water quality regulation. The success in this estimate is achieved through delimited regions that have characteristics homogeneous with the region of interest, so that the minimum flows are estimated from reliable processes. The objective of this work was to identify homogeneous regions and generate regional regression equations to estimate low flows in the Uruguai River watershed, on the state of Rio Grande do Sul state side (Brazil), using multivariate statistical analysis. The Q_{90} and Q_{95} streamflows, representing the low flows in the basin, were obtained through fluviometric stations. Easy-to-obtain geomorphological and climatic variables were used to describe the low flows in 46 sub-basins on the Uruguay River. The homogeneous regions were delimited using cluster analysis, which were confirmed from the discriminant analysis, making the process robust and reliable. Regional equations were generated for each homogeneous region and for each low flow, where the variables area, perimeter and total summer precipitation were the ones that best described the low flows. The cross-validation used confirmed the good fit of the regional equations. It is hoped that this study will allow water resource managers to make more correct decisions in processes that need to estimate low flows in Uruguai River watershed, on the Rio Grande do Sul state side, such as grant analyzes, framing water courses and charging for water use.

Regionalização hidrológica de vazões mínimas e definição de regiões homogêneas usando análise estatística multivariada na bacia hidrográfica do Rio Uruguai, lado do estado do Rio Grande do Sul, Brasil

RESUMO

Estimar vazões mínimas é fundamental para a gestão dos recursos hídricos e a regulação da qualidade da água. Para ter êxito nesta estimativa é preciso delimitar regiões que possuem características homogêneas com a região de interesse, de modo que as vazões mínimas sejam estimadas a partir de processos confiáveis. O objetivo deste trabalho foi identificar as regiões homogêneas e gerar equações de regressão regional para estimar vazões mínimas na bacia hidrográfica do Rio Uruguai, inserida no estado do Rio Grande do Sul (Brasil), usando análise estatística multivariada. As vazões Q_{90} e Q_{95} , representando as vazões mínimas na bacia, foram obtidas através de estações fluviométricas. Foram usadas variáveis geomorfológicas e climáticas de fácil obtenção para descrever as vazões mínimas em 46 sub-bacias do Rio Uruguai. As regiões homogêneas foram delimitadas usando análise de agrupamento, as quais foram confirmadas a partir da análise de discriminante, tornando o processo robusto e confiável. Foram geradas equações regionais para cada região homogênea e para cada vazão mínima, onde as variáveis área, perímetro e precipitação total de verão foram as que melhor descreveram as vazões mínimas. A validação cruzada utilizada confirmou o bom ajuste das equações regionais. Espera-se que este estudo possa permitir que gestores dos recursos hídricos tomem decisões mais corretas em processos que necessitam estimar as vazões mínimas na bacia hidrográfica do Rio Uruguai, porção do estado do Rio Grande do Sul, como análises de outorga, enquadramento dos cursos de água e cobrança pelo uso da água.

Introduction

Low flow estimation is essential for water resources management, water quality regulation and habitat protection (Smakhtin, 2001; Copatti and Copatti, 2011; Hirschmann, 2015; Schork and Zaniboni Filho, 2018; Requena et al., 2018). Traditionally, statistics have been widely used to measure the natural variability of low flows (Tsakiris et al., 2011). There are several low flow statistics that may be favorable for assessing seasonal water use and habitat needs, with the flow duration curve (FDC) being the most used (Hughes and Smakhtin, 1996; Assani et al., 2011; Mehaiguen et al., 2012; Pugliese et al., 2016; Foulon et al., 2018).

FDC is a graphical representation of the cumulative distribution of percentile flows in a watershed (Pugliesi et al., 2016) that allows to identify the flow equaled or exceeded for a given percent of time (Fouad et al., 2018). Percentile flows are used to make decisions in a variety of fields, such as environmental regulation (Over et al., 2014), transportation engineering (Dudley, 2015), and drought planning (Miller and Fox, 2017). The higher the percentile flows, the longer the flow remains in the river. Despite providing an important characterization of the streamflow behavior in a watershed, FDCs require streamflow dataset to be constructed (Lin and Wang, 2006). According to Pugliesi et al. (2016) a bigger problem occurs in watersheds ungauged, where there is a greater need to understand the streamflow behavior.

This problem can be solved using the regionalization technique. Regionalization is a useful tool to extrapolate certain hydrological information at places devoid of monitoring and without an adequate historical series dataset, using the information of regions previously monitored (WMO, 2008). Several works have already proposed this type of approach in the world (Farhan and Al-Shaikh, 2017; Fouad et al., 2018; Pagliero et al., 2019; Yang et al. 2020) and in Brazil (Pruski et al., 2012; Elesbon et al., 2015; Comini et al., 2020; Lelis et al., 2020). This technical involves two main tasks: delineating the hydrologically homogeneous regions and developing a regional estimation method (Lin and Wang, 2006). The first one is accomplished using a set of data that represent the characteristics of watersheds, which are related to low flows. These datasets are grouped according to the similarity of such characteristics. The second step is the application of a regional model of regression for each homogeneous region

that provides equations with suitable adjustment for the estimation of streamflow values.

The multivariate techniques such as cluster analysis (CA), by hierarchical methods, have been widely used in hydrology to group streamflow data and outline hydrologically homogenous regions (Chiang et al., 2002; Nosrati et al., 2015). Hierarchical methods are a well-known group of algorithms that investigates a data structure in various levels (Chiang et al., 2002; Lin and Wang, 2006; Nosrati et al., 2015; Farhan and Al-Shaik, 2017; Fouad et al., 2018). One of the hierarchical methods problems is to determine the exact number of clusters (Lin and Wang, 2006). Charrad et al. (2014) presented a variety of objective functions called cluster validity indices to identify the ideal number of clusters in a dataset. In general, the definition of hydrologically homogeneous regions is still very discussed by the scientific community, specially concerning about the uncertainty generated by its highly subjective criterion. Elesbon et al. (2015) used the inertia between jumps criterion, in which the first visible discontinuity of the graphic is defined the cutoff point, to define the homogeneous regions. However, the authors do not discuss the way these regions were validated. Beskow et al. (2016), analyzing low flow Q_{90} in Rio Grande do Sul state (Brazil), defined homogeneous regions using artificial intelligence. Clusters formed were validated using the homogeneity test, termed as H test by Hosking and Wallis (1993). Beskow et al. (2016) discuss the importance of validating homogeneous regions, diminishing the subjectivity and the uncertainties inherent to this process. Independently of the chosen method there is a lack of consistency and a certain subjectivity in the definition of the hydrologically homogeneous region (hierarchical methods) in dendrogram, and in the predefinition of the number of groups (non-hierarchical methods).

Another widely used technique of multivariate statistics is the discriminant analysis (DA). DA has been used to validate homogeneous regions, commonly generated by CA (Lin and Wang, 2006; Farhan and Al-Shaik, 2017). In addition, as an advantage over the cluster validity indices, the DA allows to explain the regional differences between homogeneous regions according to the variables used (Chiang et al., 2002). Therefore, the objective of this paper to identify the number of homogeneous regions and generate regional regression equations to estimate low flows in the Uruguai River watershed, on the

Rio Grande do Sul state side (Brazil), through using multivariate statistical analyses.

Material and Methods

The methodological development for low flow hydrological regionalization and definition of homogeneous regions in the Uruguai River watershed, on the Rio Grande do Sul state side (Brazil), consisted of using multivariate statistics, with cluster analysis (CA) coupled with discriminant analysis (DA) and multivariate regression. In the planning and management of water resources field, several studies can be found

using approximate methodologies (Byzedi et al., 2014; Youssef et al., 2015; Ahmadi et al., 2018).

The methodology makes use of the GIS (Geographic Information System) environment, in order to make the management and analysis of spatial data easier (Fraga et al., 2019). In this study, multivariate techniques were used to increase the confidence of the results obtained, especially about the definitions of homogeneous regions, followed by regression analysis to obtain regional equations. The flowchart of the methodology used is shown in Figure 1. The following topics detail the main steps for applying the methodology.

Figure 1. Flowchart of the methodology used for hydrological regionalization of low flow and definition of homogeneous regions in the Uruguai River watershed, on the Rio Grande do Sul state side, Brazil.

Study Area

The Uruguai River watershed is located in Northwestern Rio Grande do Sul state (72.6%) and southern Santa Catarina state (27.4%). This paper refers only to the analysis of the part referring to the state of Rio Grande do Sul (Figure 2), where the main uses of water resources are located (e.g. agro-industrial activities and hydroelectric plants) that depend on low flow. Another justification for working only with the Rio Grande do Sul state is due to the water use rights criteria in both states. While the Santa Catarina state uses the streamflow equal to or greater than 98% of the time (Q_{98}), the Rio Grande do Sul state uses the streamflows that

are equaled or exceeded in 90% (Q_{90}) and 95% (Q_{95}) of the time. In this work, regional equations for Q_{90} and Q_{95} were obtained.

In this study, the Uruguai River watershed encompasses two biomes. Alto Uruguai is in the Mata Atlântica biome and has original cover of Mata of Araucárias predominantly. However, along the course of the river it is observed a transition from the Mata Atlântica to the Pampa biome, where prevails the Southern fields. All the watershed is under subtropical climate, with annual thermal amplitude under 0°C along the winter and higher than 30°C in the summer.

Figure 2. Location of the Uruguai River watershed, on the Rio Grande do Sul state side, Brazil.

The use of the land and the vegetal coverage of the soil in this watershed (Rio Grande do Sul state) are classified as mixed, varying from agriculture, pasture and forest (Collischonn and Tucci, 2005). Predominant soils are the Reddish Brunizem soil and several types of Laterite, all of them with a clay texture. The North part of the Uruguai River watershed is located on the region of the Southern-Brazilian basaltic flow. According to Collischonn and Tucci (2005), that is important under the hydrological viewpoint due to the small capacity of water storage in the aquifers of this type of rock, except for the cases of great density of fractures and of sedimentary formations, located in the South part of the watershed. The Uruguai River watershed, on Rio Grande do Sul state, is in the geotectonic provinces of the Sedimentary Basin of

Paraná. According to Nakase (2008) low flows in the watershed are influenced mainly by the sedimentary formations, which increase the storage area.

The average annual precipitation is 1784mm, with an average annual temperature varying between 16 and 20°C , and an average annual evapotranspiration of 1041mm (MMA, 2006). According to Collischonn and Tucci (2005) the rainfall is relatively well distributed along the year. The natural water availability of the Uruguai River watershed is greatly influenced by important spatial and temporal variations in the rainfall regime, which is reflected in the activities developed, mainly in agriculture (MMA, 2006). The higher streamflow values occur during the period from July to October, having extreme events

in September and October, and the lower streamflow values occur in the period from December to April, with extreme events in January and March (Nakase, 2008).

The most serious cases were registered in the low middle section, where rice irrigation occurs, which showed great losses in the harvest; in the non-irrigated soy plantation areas, where the loss was even greater; and, most seriously, supply problems in the northern region of the basin, where there was serious and intense rationing (MMA, 2006). Thus, it is clear that extreme drought events, when they occur, have had their effect enhanced by the high use of water resources in the watershed, but this information has not yet been systematized, in order to obtain a history of these events, except for the Civil Defense emergency situation records (MMA, 2006). Thus, it is not possible to say with certainty what is the cause of these events, but they are one-off events, which have been assuming the tendency to become cyclical.

Data Selection Criteria

Hydrometeorological data were collected at HidroWeb - Hydrologic Information System - website, the database from the National Water Agency (ANA – Agência Nacional de Águas e Saneamento Básico). ANA is responsible for the implementation of the National System of Water Resources Management (SINGREH) in Brazil. Streamflow historical series with a collection of at least 10 years of data record (Cupak, 2017) and 31 days of lack of data for each year at the most, since these series should not exceed 10% of the lacking data (Garcia et al., 2017), were defined as database. For rainfall historical series it was adopted the criterion of at least 10 years of continuous record (Caldeira et al., 2015). Thus, 46 streamflow gauging stations and 183 rainfall gauging stations were chosen. The low flows were then calculated from the daily records, using the Q_{90} e Q_{95} Quantile.

Sub-basins for each streamflow gauging station were obtained using the Digital Elevation Model (DEM) from Shuttle Radar Topography Mission (SRTM) images with spatial resolution of 30m. The following physiographic characteristics of each sub-basin was considered in this paper: area (A, in km^2), perimeter (P, in km), and slope (S, in %). The basin centroid (X and Y centroids, in km) was used as a positioning variable, which may influence the delimitation of homogeneous regions and the potential to describe low flow regimes

(Vezza et al., 2010; Beskow et al., 2016; Fouad et al., 2018). The climatic variables total annual precipitation (p, in mm), seasonal total precipitation in summer (PS1, in mm), in autumn (PS2, in mm), in winter (PS3, in mm) and in spring (PS4, in mm), were also used as variables. The climate variables were obtained using the Thiessen Polygon's method (Schneider et al., 2017; Araújo et al., 2018). This method originates from the Voronoi diagrams (Aurenhammer, 1991) assuming that at any point in the basin the value of precipitation is equal to the weighted average of the nearest rainfall stations. In this way, it is possible to trace the areas of influence of the stations to characterize the spatial variability of precipitation (Souza et al., 2017).

The standard z-scores of all the variables were then used for the multivariate statistical analysis to lessen the effects of differences in the units used for measurements and variance and to render the data dimensionless (Gulgundi and Shetty, 2018). Then, the normality of the standardized variables was assessed by the Anderson Darling test ($p < 0.05$), as a condition for performing the multivariate analysis (Gulgundi and Shetty, 2018).

Cluster Analysis (CA)

CA is a multivariate technique that groups objects based on their characteristics, so that each object is same as the others in the cluster according to a predefined selection criterion. In this study, hierarchical method was used to form homogenous groups. This method is composed of a defined number of algorithms that verify data structure in various levels, then results are represent as a dendrogram (tree diagram) (Massart and Kaufman, 1989; Chiang et al., 2002).

In the CA, physiographic (A, P, S), positioning (X and Y centroids), climate (p, PS1, PS2, PS3, PS4) and low flows (Q_{90} , Q_{95}) variables of each sub-basin were used. Euclidean distance and Ward's method (Ward Jr., 1963) were used to identify homogeneous regions, based on various studies on low flows regionalization (Nosrati et al., 2015; Farhan and Al-Shaik, 2017). The Euclidean distance (Equation 1) shows the similarity between two samples and a distance can be represented by the difference between analytical values from the samples (Taoufik et al., 2017).

$$D_{ij} = \left[\sum_{k=1}^n (x_{ki} - x_{kj})^2 \right]^{\frac{1}{2}} \quad (1)$$

where x_{ki} and x_{kj} are observations of i -th and j -th gauging stations ($i = 1, 2, \dots, n$; and $j = 1, 2, \dots, n$), with reference to k -th variable or absolute frequency in each studied class ($k = 1, 2, \dots, n$).

In this paper, the number of cases k was represented by the number of sub-basins. The Ward's method makes use of an analysis of variance approach for evaluating the distances between clusters, in order to minimize the sum of squares of any two clusters that can be formed at each step (Lin and Wang, 2006; Tziritis et al., 2016).

The spatial variability of the sub-basins for the study area was determined from the $D_{\text{link}} / D_{\text{max}}$ ratio, which represents the relationship between the linkage distances for a particular case divided by the maximal linkage distance. To standardize the linkage distance, which is represented on the y -axis, the quotient is then multiplied by 100 (Singh et al., 2005; Gulgundi and Shetty, 2018). The length of the elements demonstrates the proximity of sub-basins. In this study, the data were grouped in a certain level on the dendrogram (Boscarello et al., 2016) using the Inertia between jumps criterion in order to define the homogeneous regions (Elesbon et al., 2015).

Discriminant Analysis (DA)

The homogeneous regions generated through CA were tested using DA, which is a supervised pattern recognition technique used for the classification of objects into exhaustive and mutually exclusive groups based on a set of independent variables. This technique is suitable when the dependent variable is a categorical variable and the independent variables are metric (Hair Jr. et al., 2009). In this study, DA was used to find the variables that discriminate between two or more groups of expected occurrences (Johnson and Wichern, 2007; Gulgundi and Shetty, 2018), generating a discriminant function (DF) for each group (Equation 2).

$$Z_{di} = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p \quad (2)$$

where Z_{di} is the vector of scores for n samples on the i -th discriminant function; variables $X_1, X_2, \dots, X_p$ are vectors for all n samples in the entire data

set; $a_{i1}, a_{i2}, \dots, a_{ip}$ are the discriminant function coefficients for the independent variables in the i -th discriminant function.

The scores for the n samples on the DF's variables have possible multiple correlations with the groups (Chiang et al., 2002). Wilk's lambda was the method used to validate the significance of DFs, which, according to Todorov (2007), it is a method used in an ANOVA (F) test of mean differences, such that the smaller the lambda index for an independent variable, the more that variable contributes to the DF. Lambda varies from 0 to 1, with 0 meaning group means differ (thus the more the variable differentiates the groups), and 1 meaning all group means are the same. The null hypothesis (H_0) was that the average population of the two groups are equal. Therefore, it is expected not to accept H_0 , because the average has to be significantly different in order to best discriminate the groups generated by the CA (Hair Jr. et al., 2009).

In this paper, DFs were constructed using the stepwise method (Hsu and Huang, 2017), with significance assessed using the F-test, with p -value < 0.05 for variables input into the function and p -value < 0.1 for variables removal, assuring that errors had a normal distribution. Subsequently, the eigenvalue matrix was generated, which is a relative measure to assess how different the groups are in the DF. I.e., the farther from 1 are the values, the bigger are the variations among the groups explained by the DF (Hair Jr. et al., 2009). Therefore, in this study, the DA tested the groups previously formed by the CA, creating a function based on the independent variables that best separate the homogenous regions.

Regional Regression Analysis and Cross-Validation

Regional regression was performed building a multi-regressive model that relates the Q_{90} and Q_{95} (dependent variable) to morphoclimatic descriptors (independent variables) as in Equation (3):

$$Q_x = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p \quad (3)$$

where x_i are the morphoclimatic descriptors and β_i are the regression coefficient.

In this study, the independent variables used to form the regional equation were chosen after selecting the variables that best characterize the homogeneous regions, i.e., after the DA result. The power regression model was chosen because it is one of the most used approaches in studies on regionalization of low flows (Lopes et al., 2016; Lujano et al., 2017). NSE_{log} , modification of the Nash-Sutcliffe coefficient (Nash and Sutcliffe, 1970), was used to quantify the performance of the adjusted models. According Razavi and Coulbay (2017), NSE_{log} is an index that best assesses low flow estimation models. The following range of values was chosen: $NSE_{log} > 0.75$, this model is considered suitable and good; $0.36 < NSE_{log} \leq 0.75$, this model is considered normal; and $NSE_{log} \leq 0.36$, this model is considered unsatisfactory.

The regional models were also assessed using the cross-validation procedure, which, according to Vezza et al. (2010), has advantages against other techniques intended for assessment of predictive errors, such as robustness and applicability to all regionalization models. The same validation process was applied by Cassalho et al. (2017). The cross-validation was analyzed for the Q_{90} and Q_{95} models using the statistics root mean square error (RMSE) and coefficient of determination (R^2), according to recommendations by Beskow et al. (2016), and the confidence index (c), proposed by Camargo and Sentelhas (1997), to allow comparison with the accuracy of the regional models. The c value was calculated using the correlation coefficient (r_{correl}) and the accuracy coefficient (d) through Equations (4) and (5). The confidence index was assessed according to the classification proposed by Camargo and Sentelhas (1997): Optimum ($c > 0.85$); Very Good ($0.76 \leq c \leq 0.85$); Good ($0.66 \leq c \leq 0.75$); Median ($0.61 \leq c \leq 0.65$); Tolerable ($0.51 \leq c \leq 0.60$); Bad ($0.41 \leq c \leq 0.50$); and Terrible ($c \leq 0.40$).

$$d = 1 - \left[\frac{\sum (Q_{est,i} - Q_{obs,i})^2}{\sum (|Q_{est,i} - Q_{obs,m}| + |Q_{obs,i} - Q_{obs,m}|)^2} \right] \quad (4)$$

where $Q_{est,i}$ are estimated Q_{90} and Q_{95} values, $Q_{obs,i}$ are estimated Q_{90} and Q_{95} values, and $Q_{obs,m}$ are the observed mean Q_{90} and Q_{95} values.

$$c = r_{correl} \cdot d \quad (5)$$

Results

Table 1 shows the independent variables for each fluvimetric gauging station, with their

respective codes, and descriptive analyzes. A great variation was observed both in size of watersheds ($349.7 - 10153.8 \text{ km}^2$) and in total annual precipitation ($671.9 - 1904.5 \text{ mm}$). That was also observed in seasonal precipitations with a variation of 54.7 to 1146.7 mm in autumn, e.g.

All variables presented normal distribution ($p < 0.05$), being able to be used in multivariate statistics analysis. CA permitted to generate a dendrogram that indicates the spatial definition pattern of the hydrologically homogeneous regions (Figure 3). The same dendrogram was generated independently of the low flow analyzed and fluvimetric stations were grouped in three similar groups.

Table 2 shows the hypothesis test, which assessed the significance of DFs for 3 and 4 homogeneous regions. In the first line of Table 2 ($N = 3$; function test = 1 to 2), the two DFs were tested together. So, it can be concluded that at least the first DF is highly significant ($p \ll 0.05$). In the second line ($N = 3$; function test = 2), only the second function was tested, concluding that it is also significant ($p < 0.05$). In relation to Wilk's lambda, the nearest of zero, the most separate are the medium values. I.e., there is a great distance between the groups (Kumar et al., 2018). Wilk's lambda also demonstrated that the first function is the one that has the best discrimination power, since its value is the one that is nearest of zero. The values of Chi-square confirmed the significance of the DFs. I.e., how good each function separates the observation into the groups. Then, it was accomplished a test with four homogeneous regions. It can be observed in the last line of Table 2 ($N = 4$; function test = 3) that the value of Wilks' lambda was higher, indicating that the formation of three regions is the best configuration for the study site.

As defined in the DA that the best configuration for the study area would be three homogeneous regions, two discriminating functions were generated. The first DF presents a percentage of 72.8%. That is the function that best contributes to demonstrate the difference between the groups.

Table 1. Values of independent variables used in the study and descriptive analyzes.

n	Code	A* (km ²)	P* (km)	Centroid*		S* (%)	p* (mm)	PS1* (mm)	PS2* (mm)	PS3* (mm)	PS4* (mm)
				X (km)	Y (km)						
1	72400000	1573.6	206.6	442.2	6888.5	9.5	1807.1	147.6	131.3	166.6	157.2
2	72430000	1737.8	218.2	441.5	6890.7	9.9	1815.3	147.8	146.3	184.0	174.7
3	72580000	1159.7	170.3	387.9	6889.3	9.7	1599.3	133.5	110.3	152.2	139.0
4	72630000	2993.3	298.8	398.8	6898.6	13.2	1723.0	140.8	206.7	158.4	152.5
5	72680000	3907.7	354.6	398.3	6907.9	14.2	1744.2	142.6	125.4	157.3	155.7
6	73480000	3965.4	370.8	347.7	6922.3	10.3	1769.0	141.8	127.7	162.3	343.2
7	74205000	466.2	106.6	314.5	6915.4	11.6	1728.4	142.9	125.5	155.4	152.1
8	74210000	2746.3	327.9	309.8	6892.5	9.0	1813.1	141.1	140.2	156.6	165.4
9	74270000	6223.9	468.8	301.2	6918.6	12.2	1840.1	143.5	146.9	155.2	211.3
10	74370000	2153.9	257.9	255.1	6937.0	10.4	671.9	52.4	54.7	58.1	58.9
11	74460000	536.0	126.4	241.2	6906.9	7.1	1904.5	144.9	161.1	152.1	176.8
12	74470000	1639.2	260.0	232.4	6930.9	9.6	1830.1	140.6	150.3	147.5	171.4
13	74600000	2407.7	245.6	793.9	6920.9	9.4	1781.9	136.0	144.7	145.2	164.4
14	74610000	2407.6	245.3	793.8	6920.8	9.3	1781.8	135.9	144.6	145.3	164.3
15	74700000	1216.9	196.7	767.8	6911.9	7.3	1805.4	136.3	150.0	144.5	165.6
16	74750000	349.7	96.4	749.0	6907.2	7.3	1827.8	138.9	156.4	146.0	168.0
17	74880000	878.2	181.9	749.1	6890.4	6.9	1890.8	141.6	160.6	143.4	183.3
18	74900000	1346.7	268.0	736.1	6893.5	7.2	1885.0	139.5	159.2	136.3	189.4
19	75155000	2080.6	242.6	257.4	6872.7	7.9	1723.7	135.3	1146.7	156.2	171.8
20	75185000	655.2	156.6	238.7	6846.2	6.9	1705.1	130.0	1137.9	149.5	154.9
21	75186000	655.9	156.7	238.7	6846.2	6.9	1710.9	138.6	395.0	125.3	152.0
22	75200000	867.0	156.7	225.2	6833.9	6.6	1856.0	142.2	814.0	745.9	170.2
23	75205000	1039.2	174.7	221.6	6836.4	6.5	1856.5	142.3	939.4	860.2	170.7
24	75230000	5500.5	446.2	232.4	6863.1	7.3	1801.2	137.6	416.4	393.1	267.4
25	75295000	2321.6	312.8	791.9	6817.9	5.9	1902.3	143.0	446.3	418.8	174.2
26	75300000	9573.0	595.0	803.5	6851.5	7.0	1820.2	138.5	394.4	375.2	230.3
27	75320000	10153.8	625.0	799.1	6852.3	6.9	1885.1	140.1	198.8	188.9	185.8
28	75400000	1003.5	155.3	762.5	6816.3	6.1	1887.7	144.0	156.6	144.7	179.3
29	75430000	957.6	155.8	739.9	6799.3	5.7	1470.7	114.8	126.5	115.7	133.2
30	75450000	3474.6	317.1	741.0	6817.7	5.7	1806.7	138.5	152.0	143.0	167.3
31	75500000	5626.0	478.4	720.8	6832.2	5.4	1795.9	141.1	155.2	139.7	161.9
32	75600000	406.0	109.1	719.0	6796.7	5.8	1761.5	137.2	147.6	145.0	157.4
33	75700000	3995.9	340.0	680.8	6809.0	4.7	1807.6	636.9	160.9	136.0	160.9
34	76085000	1668.0	212.1	803.3	6755.8	6.6	1712.9	132.3	147.6	141.2	149.8
35	76100000	3028.5	299.1	790.1	6750.4	9.4	1742.9	136.9	149.4	142.4	152.2
36	76120000	3590.6	375.8	783.1	6746.4	9.3	1724.6	135.6	146.7	141.7	150.8
37	76360001	1260.5	235.8	742.4	6660.0	4.9	1555.0	115.3	125.3	131.7	146.1
38	76380000	1986.7	301.1	735.3	6665.7	5.0	1565.4	117.7	126.4	130.9	146.9
39	76440000	2514.1	298.8	758.2	6767.7	7.7	1879.3	148.5	155.8	147.0	169.6
40	76460000	996.3	166.5	725.1	6770.6	7.6	1812.8	146.0	154.8	140.7	162.9
41	76490000	4965.2	447.9	737.9	6758.5	7.9	1822.2	145.5	152.5	142.8	163.9
42	76550000	408.2	113.3	675.4	6747.4	7.8	1707.6	141.8	152.7	124.1	150.7
43	76650000	2774.9	323.2	663.2	6770.1	6.6	1646.4	144.2	146.0	122.2	136.4
44	76742000	1273.4	221.7	653.7	6660.7	7.8	1530.7	122.7	133.1	106.7	147.6
45	76750000	6503.8	457.8	629.8	6650.3	6.4	1477.0	121.5	125.9	103.5	141.4
46	79400000	1367.0	206.7	778.5	6521.3	4.0	1357.1	106.0	115.8	120.6	110.0
Minimum		349.7	96.4	221.6	6521.3	4.0	671.9	52.4	54.7	58.1	58.9
Maximum		10153.8	625.0	803.5	6937.0	14.2	1904.5	636.9	1146.7	860.2	343.2

Table 1. To be continue.

Average	2573.0	271.4	571.9	6827.4	7.8	1729.2	146.1	242.6	186.9	166.5
SD*	2247.7	126.0	225.0	91.2	2.2	202.1	75.6	255.6	149.1	39.6

* A – Area; P – Perimeter; X – Centroid X; Y – Centroid Y; S – Slope; p – Annual total precipitation; PS1 – Summer total precipitation; PS2 – Autumn total precipitation; PS3 – Winter total precipitation; PS4 – Spring total precipitation; SD – standard deviation.

Figure 3. Hydrologically homogeneous regions generated by Cluster Analysis.

Table 2. Wilk's lambda for 3 and 4 homogeneous regions (N) generated by CA.

N	Function test	Wilk's lambda	Chi-square	df	Sig.
3	1 to 2	0.091	63.575	16	0.000
	2	0.422	22.839	7	0.002
4	1 to 2	0.033	95.608	12	0.000
	2 to 3	0.148	53.525	6	0.000
	3	0.489	20.052	2	0.000

Figure 4 presents the spatial distribution of the three hydrologically homogeneous regions generated by the CA and confirmed using the DA in the study area. After the gradual input and the removal of some variables, the variables were reduced from 10 to 5 (A, P, X, Y, and PS1). Therefore, the most important variables in the definition of homogeneous regions in the study area. This result indicates that other variables were

not significant ($p > 0,05$). i.g., it is possible to reduce the quantity of variables to be used in the regional regression without interfering in the formation of homogeneous regions. Results pointed the importance of the effects of seasonality on the separation of homogeneous regions, since the variable PS1 was present in the formation of the DFs (Equations 6 and 7).

$$Z_1 = 0.734A + 0.189P - 0.873X + 0.773Y + 0.307PS1 \quad (6)$$

$$Z_2 = -1.083A + 0.306P - 0.076X + 0.059Y + 0.631PS1 \quad (7)$$

Figure 4. Spatial distribution of the three homogeneous regions generated for the Uruguai River watershed.

All equations generated to estimate the Q_{90} and Q_{95} streamflows were considered “suitable and good”, according to the NSE_{log} index, indicating that the functions adjustment was satisfactory (Table 3). Cross-validation showed that the two regional equations generated for region 1 were considered “very good”, according to the c index. The other equations (region 2 and region 3) were considered “good” to estimate the flows Q_{90} and Q_{95} in the Uruguai River watershed, on Rio Grande do Sul state side. The estimation error RMSE ranged from $1.73 \text{ m}^3 \cdot \text{s}^{-1}$ (Q_{95} region 1) to $3.70 \text{ m}^3 \cdot \text{s}^{-1}$ (Q_{90} region 3), being considered satisfactory in this study.

In the Table 4 is presented the limit of application of the variables of each equation. Results demonstrate a great regularity since only three variables generate all the equations (A – area; P – perimeter; and $PS1$ –summer total precipitation). Region 1 presented the best adjustment only with the variable perimeter; region

2 presented the best adjustment with the variables area and perimeter; and region 3 presented the best adjustment with the variables area and summer total precipitation.

Low flow quantiles regionally estimated by multivariate statistical analysis are visually compared with observed low flow quantiles in Figure 5 and Figure 6. These Q-Q plots show regionally estimated quantiles to be similar to observed quantiles, although several exceptions exist. Low flow quantiles in the region 1 show one clear outlier, site “74370000”. The reason for their low flow underestimation was attributed to having a very low perimeter value compared with sites of a similar size. Low flow quantiles in the region 3 show two clear outlier, site “75500000” and “76650000”. In this case, low $PS1$ values were responsible for underestimating the low flow quantiles.

Table 3. Regional equations to estimate Q_{90} and Q_{95} obtained in this study.

FDC _a	Region	Regional Regression Equation ^b	Fuction adjustment			Cross-validation			
			NSE _{log}	Effectiveness	RMSE ^a	r _{correl}	d	c	Performa nce
Q ₉₀	1	$1.15 \times 10^{-3} \times (P^{1.63})$	0.97	Suitable and Good	2.08	0.87	0.98	0.85	Very Good
	2	$1.39 \times 10^{-10} \times (A^{-7.55} \times P^{14.95})$	0.96	Suitable and Good	3.55	0.86	0.85	0.73	Good
	3	$3.43 \times 10^{-14} \times (A^{0.71} \times PS1^{5.53})$	0.88	Suitable and Good	3.70	0.80	0.92	0.74	Good
Q ₉₅	1	$1.22 \times 10^{-3} \times (P^{1.57})$	0.95	Suitable and Good	1.73	0.82	0.98	0.80	Very Good
	2	$7.58 \times 10^{-11} \times (A^{-7.69} \times P^{15.19})$	0.95	Suitable and Good	2.75	0.85	0.85	0.72	Good
	3	$6.74 \times 10^{-15} \times (A^{0.70} \times PS1^{5.81})$	0.77	Suitable and Good	3.38	0.76	0.87	0.66	Good

^aFDC and REMQ, in $m^3 \cdot s^{-1}$; ^bA, in km^2 ; P, in km; and PS1, in mm.

Table 4. Application limits of regional regression equations in the study area.

Region	n*	P (km)	A (km^2)	PS1 (mm)
1	10	106.6 - 327.9		
2	12	156.6 - 625.0	867.0 - 10153.8	
3	24		349.7 - 6503.8	106.0 - 148.5

* n is the number of streamflow gauging stations.

Discussion

This paper proposed to carry out the low flow hydrological regionalization in the Uruguai River watershed, side of the Rio Grande do Sul state (Brazil), using cluster analysis (CA) and discriminant analysis (DA). Although this work presents a case study in Brazil, the methodology used can be applied in any watershed. The use of robust multivariate techniques to define homogeneous regions is encouraged in order to reduce the uncertainties generated by the regression models.

We chose to analyze morphoclimatic variables as descriptors of low flows, since they are easy to obtain, in order to subsidize grant processes, an instrument used by the Brazilian legislation when planning the use of the water

resources (Brasil, 1997). Similar proposal was accomplished by Beskow et al. (2016), when they analyzed the Q_{90} for the Rio Grande do Sul state using the drainage area of hydrographic sub-basins. Incorporating other variables in the study may give robustness to regionalization models. That was done by Pruski et al. (2012), in a study accomplished in the Pará River watershed, sub-basin of São Francisco River, Brazil. The authors added the annual average precipitation climatic variable in the study of regionalization of low flows. The precipitation that occurs in the drainage area of a watershed is a factor that interferes directly on streamflows behavior, that is the reason why inclusion of precipitation as an explanatory variable may represent a great improvement in the streamflow regionalization model (Comini et al., 2020; Yang et al., 2020).

Figure 5. Co-relation between observed and estimated Q_{90} streamflow. Particular fluviometric stations are identified.

Analyzing low flows in watershed is a challenge because the streamflow is controlled by base flows that may arise beyond the superficial outlines of the ground (Rumsey et al., 2015). The Uruguai River watershed is in the borders of the states of Rio Grande do Sul and Santa Catarina, besides being in the frontier of Argentina. That increases the complexity of this study since the base flow in the basin can come from those places. We emphasize that studies with this complexity was not done in that region yet, what is a future tendency. In this study, we considered only the limits of the Rio Grande do Sul state as an analysis of the regionalization study and, in the future, it may be used for water use rights criteria analysis,

Figure 6. Co-relation between observed and estimated Q_{95} streamflow. Particular fluviometric stations are identified.

similar to what was considered by Beskow et al. (2016) and Farhan and Al-Shaikh (2017).

Other important point to be considered is the location of fluviometric stations. In studies of this kind, which propose to study low flows, it is important to consider the natural streamflows, without the interference of dams. However, the presence of dams is remarkable in this watershed. It was used all the fluviometric stations available by the ANA, possible to be used after passing by the statistic criteria. We understand that the number of stations is not the ideal, but that is the reality of several watershed poorly monitored in Brazil (Lelis et al., 2020).

Results show that the best configuration of the hydrologically homogeneous regions of the Uruguai River watershed, on Rio Grande do Sul state side, was the separation in three regions. There is not an agreement in literature about the ideal number of groups, being a particular characteristic of each region. Modarres (2010) affirms that for management purposes it is recommended a greater number of groups to interpret the characteristics of the regions and to use them with more confidence in the applicability of the regional equations. However, Vezza et al. (2010) suggest that the possible consideration of only one homogeneous region is the simplest way for executing the regionalization. The same researchers pointed that the hypothesis of applying a general model generalize all the different important processes for the analysis of low flows. That is why it is important to have mechanisms to verify if the formation of groups is suitable with the particularities of the drainage basin. In this study it was used the DA and the results showed that the three regions formed by the CA are the best configuration for the study site.

Models of regional regression indicated that the area (A), perimeter (P) and summer total precipitation (PS1) were the variables that best distinguished the groups formed. The presence of the variable area in 4 of the 6 equations regional agrees with several published studies about regionalization of low flows. Although the variable area did not influence some regional equations, somehow it was represented by the variable perimeter. According to Gasques et al. (2018), the drainage area has a good correlation with the other physical characteristics of the basin and presents influence on water availability along the hydrography. In the comprehensive review of Razavi and Coulibaly (2013), the drainage area is one of the most used attributes and it is one that results in more satisfactory results of the models. Lopes et al. (2016) applied as independent variables: annual average precipitation, drainage density and area. The authors obtained the best performance of the model with only the variable area as independent variable. Beskow et al. (2016) obtained results for regionalization of Q_{90} for the Uruguai River watershed using only the drainage area as an explanatory variable. The authors found reliability indexes c lower than the ones found in this study, demonstrating that the methodology used here is satisfactory and that other explanatory

variables may interfere in the estimative of low flows.

Results point that the use of climatic variables in low flows regionalization models improves equations performances, as it can be observed with the input of the variable summer seasonal precipitation (PS1). This variable collaborated to provide the best adjustment in the regional equations of region 3. The lower flow values in the watershed occurs in the summer period, as described by Nakase (2008). That justifies the results found. Besides, region 3 is where sedimentary geological formations are located, which contributes to the formation of the base flow. Similar results were found by Trudel et al. (2016) and Requena et al. (2018). Perimeter is the main variable that characterizes region 1, appearing also in region 2. Both regions are located mainly over crystalline shields, where water infiltration in order to generate the base flows is lower. Thus, area and perimeter were highlighted when compared to PS1. However, this is a hypothesis that will be analyzed in the future since perimeter may have appeared in the results with a greater importance due to the presence of the variable area.

In this study, only the morphoclimatic variables were analyzed in order to evaluate the low flows in the Uruguai River watershed. We emphasize that further studies should approach variables that may best characterize this environment, such as, lithology (percentage of permeable formation), vegetation coverage and index of use and occupation of the land. It is important to emphasize that the use of inadequate regional equations can result in the estimation of unreliable low flows. Those distortions may occur due to a series of errors, such as the unsatisfactory choose of the independent variables as well as the errors inherent to the methodology. Such factors may reveal spurious estimation of low flows, which may impact on studies related to the policy and the management of water resources.

Conclusions

Based on the results presented and discussed, it can be concluded that:

- (i) The methodology used, based on multivariate statistical analyses, proved to be robust and suitable for individualization of homogeneous regions, indicating regional

low flow equations to be used by water resources management bodies in the Uruguai River watershed, on Rio Grande do Sul state side, Brazil.

- (ii) After performing the CA, the DA increased confidence in defining the three hydrologically homogeneous regions aiming at the regionalization of the Q_{90} and Q_{95} streamflows in the Uruguai River watershed.
- (iii) The variables area, perimeter and total summer precipitation are responsible for the best performance of adjustment of the regional regression equations in order to estimate the Q_{90} and Q_{95} streamflows concerning the 46 sub-basins in the the Uruguai River watershed.
- (iv) The adjusted regional equations performed particularly well in estimating Q_{90} and Q_{95} streamflows, requiring few variables. This is due to the robust methodology for defining homogeneous regions and cross-validation used in this study. Therefore, the equations proposed can be used in the Uruguai River watershed, on the Rio Grande do Sul state side (Brazil), to estimate the Q_{90} and Q_{95} streamflows in sites ungauged.

Acknowledgments

The authors would like to thank the Brazilian' Agencie CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico - National Council for Scientific and Technological Development) for its financial support during the research development.

References

Ahmadi, F., Radmaneh, F., Sharifi, M.R., Mirabbasi, R., 2018. Bivariate frequency analysis of low flow using copula functions (case study: Dez River Basin, Iran). *Environmental Earth Sciences* [online] 77. Available: <https://doi.org/10.1007/s12665-018-7819-2>. Access: 01 sept. 2019.

Araújo, F.C., Mello, E.L., Gollin, G.M., Quadros, L.E., Gomes, B.M., 2018. Streamflow regionalization in Piquiri river basin. *Engenharia Agrícola* 38, 22-31.

Assani, A.A., Chalifour, A., Légraré, G., Manouane, C-S., Leroux, D., 2011. Temporal Regionalization of 7-Day Low Flows in the St.

Lawrence Watershed in Quebec (Canada). *Water Resources Management* 25, 3559-3574.

Aurenhammer, F., 1991. Voronoi diagrams - a survey of fundamental geometric data structure. *ACM Computing Surveys* 23, 345-405.

Beskow, S., Mello, C.R., Vargas, M.M., Corrêa, L.L., Caldeira, T.L., Durães, M.F., Aguiar, M.S. de, 2016. Artificial intelligence techniques coupled with seasonality measures for hydrological regionalization of Q_{90} under Brazilian conditions. *Journal of Hydrology* 541, 1406-1419.

Boscarello, L., Ravazzani, G., Cislighi, A., Mancini, M., 2016. Regionalization of Flow-duration Curves through Catchment Classification with Streamflow Signatures and Physiographic–Climate Indices. *Journal of Hydrologic Engineering* 21, 1-17.

Brasil, 1997. Lei nº 9.433, de 8 de janeiro de 1997 - Institui a Política Nacional de Recursos Hídricos, cria o Sistema Nacional de Gerenciamento de Recursos Hídricos, regulamenta o inciso XIX do art. 21 da Constituição Federal, e altera o art. 1º da Lei nº 8.001, de 1, 1997 [online]. Available: http://www.planalto.gov.br/ccivil_03/LEIS/L9433.htm. Access: 20 jun. 2020.

Byzedi, M., Saghaffian, B., Mohammadi, K., Siosemarde, M., 2014. Regional analysis of streamflow drought: a case study in southwestern Iran. *Environmental Earth Sciences* 71, 2955-2972.

Caldeira, T.L., Beskow, S., Mello, C.R., Faria, L.C., Souza, M.R., Guedes, H.A.S., 2015. Modelagem probabilística de eventos de precipitação extrema no estado do Rio Grande do Sul. *Revista Brasileira de Engenharia Agrícola e Ambiental* 19, 197-203.

Camargo, A.P., Sentelhas, P.C., 1997. Avaliação do desempenho de diferentes métodos de estimativas da evapotranspiração potencial no Estado de São Paulo, Brasil. *Revista Brasileira de Agrometeorologia* 5, 89 – 97.

Cassalho, F., Beskow, S., Vargas, M.M., Moura, M.M., Ávila, L.F., Mello, C.R., 2017. Hydrological regionalization of maximum stream flows using an approach based on L-moments. *Revista Brasileira de Recursos Hídricos* [online] 22. Available: <http://dx.doi.org/10.1590/2318-0331.021720160064>. Access: 10 oct. 2019.

- Charrad, M., Ghazzali, N., Boiteau, V., Niknafs, A., 2014. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software* 61, 1-36.
- Chiang, S-M., Tsay, T-K., Nix, S.J., 2002. Hydrologic Regionalization of Watersheds. I: Methodology Development. *Journal of Water Resources Planning and Management* 128, 3-11.
- Collischonn, W., Tucci, C.E.M., 2005. Previsão Sazonal de Vazão na Bacia do Rio Uruguai 1: Ajuste e Verificação do Modelo Hidrológico Distribuído. *Revista Brasileira de Recursos Hídricos* 10, 43-59.
- Comini, U.B., Silva, D.D., Moreira, M.C., Pruski, F.F., 2020. Hydrological modelling in small ungauged catchments. *Anais da Academia Brasileira de Ciências* 92, 1-16.
- Copatti, C.E., Copatti, B.R., 2011. Seasonal variation and diversity of fishes of the Camará River, Uruguay Basin. *Biota Neotropica* 11, 265-271.
- Cupak, A., 2017. Initial results of nonhierarchical cluster methods use for low flow Grouping. *Journal of Ecological Engineering* 18, 44-50.
- Dudley, R.W., 2015. Regression Equations for Monthly and Annual Mean and Selected Percentile Streamflows for Ungaged Rivers in Maine. US Geological Survey Scientific Investigations Report 2015-5151 [online]. Available: <https://doi.org/10.3133/sir20155151>. Access: 10 jan. 2020.
- Elesbon, A.A.A., Silva, D.D., Sediya, G.C., Guedes, H.A.S., Ribeiro, C.A.A.S., Ribeiro, C.B.M., 2015. Multivariate statistical analysis to support the minimum streamflow regionalization. *Engenharia Agrícola* 35, 838-851.
- Farhan, Y., Al-Shaikh, N., 2017. Quantitative Regionalization of W. Mujib-Wala Sub-Watersheds (Southern Jordan) Using GIS and Multivariate Statistical Techniques. *Open Journal of Modern Hydrology* 7, 165-199.
- Fouad, G., Skupin, A., Tague, C.L., 2018. Regional regression models of percentile flows for the contiguous United States: Expert versus data-driven independent variable selection. *Journal of Hydrology: Regional Studies* 17, 64-82.
- Foulon, E., Rousseau, A.N., Gagnon, P., 2018. Development of a methodology to assess future trends in low flows at the watershed scale using solely climate data. *Journal of Hydrology* 557, 774-790.
- Fraga, M.S., Silva, D.D., Elesbon, A.A.A., Guedes, H.A.S., 2019. Methodological proposal for the allocation of water quality monitoring stations using strategic decision analysis. *Environmental Monitoring and Assessment* 191, 776-795.
- Garcia, F., Folton, N., Oudin, L., 2017. Which objective function to calibrate rainfall-runoff models for low-flow index simulations?. *Hydrological Sciences Journal* 62, 1149-1166.
- Gasques, A.C.F., Neves, G.L., Santos, J.D., Mauad, F.F., Okawa, C.M.P., 2018. Regionalização de vazões mínimas: breve revisão teórica. *Revista Eletrônica de Engenharia Civil* 14, 60-70.
- Gulgundi, M.S., Shetty, A., 2018. Groundwater quality assessment of urban Bengaluru using multivariate statistical techniques. *Applied Water Science* 8, 1-15.
- Hair Jr, J.F., Black, W.C., Babin, B.J., Anderson, R.E., Tatham, R.L., Sant'Anna, M.A.G.A.S., 2009. *Análise multivariada de dados*. Porto Alegre: Bookman.
- Hirschmann, A., 2015. Variação da composição e estrutura da ictiofauna em três rios com e sem influência de barramento no sul do Brasil. *Neotropical Biological Conservation* 10, 143-151.
- Hosking, J.R.M., Wallis, J.R., 1993. Some statistics useful in regional frequency analysis. *Water Resources Research* 29, 271 – 281.
- Hsu, N-S., Huang, C-J., 2017. Estimation of Flow Duration Curve at Ungauged Locations in Taiwan. *Journal of Hydrologic Engineering* [online] 22. Available: [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001511](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001511). Access: 12 may 2019.
- Hughes, D.A., Smakhtin, V., 1996. Daily flow time series patching or extension: a spatial interpolation approach based on flow duration curves. *Hydrological Sciences Journal* 41, 851-871.
- Johnson, R.A., Wichern, D.W., 2007. *Applied multivariate statistical analysis*. New Jersey: Prentice-Hall.
- Kumar, B., Singh, U.K., Ojha, S.N., 2018. Evaluation of geochemical data of Yamuna River using WQI and multivariate statistical analyses: a case study. *International Journal of River Basin Management* 17, 143-155.

- Lelis, L.C.S., Nascimento, J.G., Duarte, S.N., Pacheco, A.B., Bosquilia, R.W.D., Wolff, W., 2020. Assessment of hydrological regionalization methodologies for the upper Jaguari River basin. *Journal of South American Earth Sciences* 97, 1-10.
- Lin, G-F., Wang, C-M., 2006. Performing cluster analysis and discrimination analysis of hydrological factors in one step. *Advances in Water Resources* 29, 1573-1585.
- Lopes, T.R., Prado, G., Zolin, C.A., Paulino, J., Antoniel, L.S., 2016. Regionalização de Vazões Máximas e Mínimas para a Bacia do Rio Ivai – PR. *Irriga* 21, 188-201.
- Laura, A.L., Aragon, J.P.Q., Laura, E.L., 2017. Regionalización de Caudales Mensuales en la Región Hidrográfica del Titicaca, Perú. *Revista de Investigaciones Altoandinas* 19, 219-230.
- Massart, D.L., Kaufman, L., 1989. The interpretation of analytical data by the use of cluster analysis. New York: Wiley.
- Mehaiguen, M., Meddi, M., Longobardi, A., Toumi, S., 2012. Low flows quantification and regionalization in North West Algeria. *Journal of Arid Environments* 87, 67-76.
- Miller, R.B., Fox, G.A., 2017. A tool for drought planning in Oklahoma: estimating and using drought-influenced flow exceedance curves. *Journal of Hydrology: Regional Studies* 10, 35-46.
- MMA – Ministério do Meio Ambiente, 2006. Caderno da Região Hidrográfica do Uruguai. Secretaria de Recursos Hídricos. Brasília (Brasil): Ministério do Meio Ambiente.
- Modarres, R., 2010. Regional dry spells frequency analysis by L-Moment and multivariate analysis. *Water Resources Management* 24, 2365-2380.
- Nakase, L.M., 2008. Plano de Desenvolvimento Sustentável da Região da Bacia do Rio Uruguai. Componente 1 – Diagnóstico da Região da Bacia do Rio Uruguai. Relatório Final. Cooperação Técnica BID ATN/JC-9952-BR [online]. Available: http://www.mpf.mp.br/atuacao-tematica/ccr4/dados-da-atuacao/informes/pdfs/Relatorio_Diagnostico_Versao%20Final.pdf. Access: 17 may 2020.
- Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models: Part I – a discussion of principles. *Journal of Hydrology* 10, 282-290.
- Nosrati, K., Laaha, G., Sharifnia, S.A., Rahimi, M., 2015. Regional low flow analysis in Sefidrood Drainage Basin, Iran using principal component regression. *Hydrology Research* 46, 121-135.
- Over, T.M., Riley, J.D., Sharpe, J.B., Arvin, D.V., 2014. Estimation of Regional Flow-Duration Curves for Indiana and Illinois. US Geological Survey Scientific Investigations Report 2014-5177 [online]. Available: <https://doi.org/10.3133/sir20145177>. Access: 15 jan. 2020.
- Pagliero, L., Bouraoui, F., Diels, J., Willems, P., McIntyre, N., 2019. Investigating regionalization techniques for large-scale hydrological modelling. *Journal of Hydrology* 570, 220-235.
- Pruski, F.F., Nunes, A.A., Rego, F.S., Souza, M.F., 2012. Extrapolação de equações de regionalização de vazões mínimas: Alternativas para atenuar os riscos. *Water Resources and Irrigation Management* 1, 51-59.
- Pugliesi, A., Farmer, W.H., Castellarin, A., Archfield, S.A., Vogel, R.M., 2016. Regional flow duration curves: Geostatistical techniques versus multivariate regression. *Advances in Water Resources* 96, 11-22.
- Razavi, T., Coulibaly, P., 2013. Streamflow prediction in ungauged basins: review of regionalization methods. *Journal of Hydrologic Engineering* 18, 958-975.
- Razavi, T., Coulibaly, P., 2017. An evaluation of regionalization and watershed classification schemes for continuous daily streamflow prediction in ungauged watersheds. *Canadian Water Resources Journal* 42, 2-20.
- Requena, A.I., Ouarda, T.B.M.J., Chebana, F., 2018. Low-flow frequency analysis at ungauged sites based on regionally estimated streamflow. *Journal of Hydrology* 563, 523-532.
- Rumsey, C.A., Miller, M.P., Susong, D.D., Tillman, F.D., Anning, D.W., 2015. Regional scale estimates of baseflow and factors influencing baseflow in the Upper Colorado River Basin. *Journal of Hydrology: Regional Studies* 4, 91-107.
- Schneider, E.H.M., Barbosa, A.G., Rocha, I.M.L., Mendes, L.A., 2017. Regionalização das vazões de permanência para as bacias hidrográficas do Estado de Sergipe segundo

- suas regiões climáticas. *Scientia Plena* 13, 1-7.
- Schork, G., Zaniboni Filho, E., 2018. Influence of Spatial Gradient Caused by a Large Dam on Fish Assemblage in a Subtropical Reservoir – Upper Uruguay River. *Boletim do Instituto de Pesca* 44, 1-11.
- Singh, K.P., Malik, A., Sinha, S., 2005. Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques: a case study. *Analytica Chimica Acta* 538, 355–374.
- Smakhtin, V.U., 2001. Low flow hydrology: a review. *Journal of Hydrology* 240, 147-186.
- Souza, N.S., Souza, W.J., Cardoso, J.M.S., 2017. Caracterização hidrológica e influência da cobertura do solo nos parâmetros de vazão do Rio das Fêmeas. *Engenharia Sanitária e Ambiental* 22, 453-462.
- Taoufik, G., Khouni, I., Ghrabi, A., 2017. Assessment of physico-chemical and microbiological surface water quality using multivariate statistical techniques: a case study of the Wadi El-Bey river, Tunisia. *Arabian Journal of Geosciences* [online] 10. Available: <https://doi.org/10.1007/s12517-017-2898-z>. Access: 13 sep. 2019.
- Todorov, V., 2007. Robust selection of variables in linear discriminant analysis. *Statistical Methods and Applications* 15, 395-407.
- Trudel, M., Doucet-Généreux, P-L., Leconte, R., Côté, B., 2016. Vulnerability of water demand and aquatic habitat in the context of climate change and analysis of a no-regrets adaptation strategy: Study of the Yamaska River Basin, Canada. *Journal of Hydrologic Engineering* [online] 21. Available: [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001298](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001298). Access: 15 oct. 2019.
- Tsakiris, G., Nalbantis, I., Cavadias, G., 2011. Regionalization of low flows based on canonical correlation Analysis. *Advances in Water Resources* 34, 865-872.
- Tziritis, E., Skordas, K., Kelepertsis, A., 2016. The use of hydrogeochemical analyses and multivariate statistics for the characterization of groundwater resources in a complex aquifer system. A case study in Amyros river basin, Thessaly, central Greece. *Environmental Earth Sciences* 75, 1–11.
- Vezza, P., Comoglio, C., Rosso, M., Viglione, A., 2010. Low flows regionalization in North-Western Italy. *Water Resources Management* 24, 4049-4074.
- Ward Jr, J.H., 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* 58, 236-244. (Published online: 10 Apr. 2012).
- WMO, 2008. Manual on Low-Flow Estimation and Prediction. Operational Hydrology Report n°50. World Meteorological Organization Germany [online]. Available: http://www.wmo.int/pages/prog/hwrr/publications/low-flow_estimation_prediction/WMO%201029%20en.pdf. Access: 20 jun. 2020.
- Yang, X., Magnusson, J., Huang, S., Beldring, S., Xu, C-Y., 2020. Dependence of regionalization methods on the complexity of hydrological models in multiple climatic regions. *Journal of Hydrology* [online] 582. Available: <https://doi.org/10.1016/j.jhydrol.2019.124357>. Access: 08 aug. 2020.
- Youssef, A.M., Pradhan, B., Jebur, M.N., El-Harbi, H.M., 2015. Landslide susceptibility mapping using ensemble bivariate and multivariate statistical models in Fayfa area, Saudi Arabia. *Environmental Earth Sciences* 73, 3745-3761.