

An Acad Bras Cienc (2021) 93(1): e20190747 DOI 10.1590/0001-3765202120190747 Anais da Academia Brasileira de Ciências | Annals of the Brazilian Academy of Sciences Printed ISSN 0001-3765 | Online ISSN 1678-2690 www.scielo.br/aabc | www.fb.com/aabcjournal

ENGINEERING SCIENCES

Regionalization of flow duration curves in the Amazon with the definition of homogeneous regions via fuzzy C-means

FRANCISCO C.L. PESSOA, CLAUDIO J.C. BLANCO & EVANICE P. GOMES

Abstract: Data insufficiency is one of the main challenges faced in hydrological studies, including a lack of knowledge regarding flow duration curves (FDCs). Thus, homogeneous regions of streamflow were identified in the Amazon using the Fuzzy C-Means (FCM) method. The PBM index was used to validate the clustering obtained via FCM, in turn, a homogeneity test based on the L-moment was applied to confirm the homogeneity in each defined region. Linear, power, exponential, logarithmic, quadratic and cubic mathematical models were fitted to the FDCs observed in the homogeneous regions. The models are the result of multiple regression analyses involving the parameters of the fitted FDC and the physico-climatic characteristics of the watersheds. The models were validated using the Jack-knife cross-validation method. The validation was satisfactory, with NASH coefficients higher than 0.50. Additionally, the standard deviation (RSR) of observations was less than 0.70, and the averages of the relative mean square error did not exceed 12.26%. These results are relevant for 89.91% of the analyzed watersheds and 73.58% of the study area. Thus, FDCs may be estimated in large parts of the Amazon, thereby making the methodology presented a valuable tool to support projects involving the planning and management of water resources.

Key words: Clusters, multiple regression, PBM index, regionalization.

INTRODUCTION

One of the conditions for the proper planning of water resources is knowledge of the behavior of rivers and their streamflow regimes. This knowledge requires the continuous collection and interpretation of the data obtained at streamflow gauge stations because reliability becomes more important as the historical series become more extensive. However, the lack of gauged sites and the limited availability of streamflow observations creates a relevant problem for the estimation of flow duration curves (FDCs) that is usually solved through the application of regionalization models (Castellarin et al. 2007, Ganora et al. 2009). According to Li et al. (2010), hydrological regionalization is a technique that allows the estimation of variables such as rainfall and streamflow, and of hydrological functions, such as FDCs. It is necessary to know the hydrological processes and to understand the spatiotemporal heterogeneity of the morphoclimatic properties of the watersheds.

These functions serve as analytical tools for hydrological and environmental problems related to the use of the water in a watershed for various purposes, such as hydroelectric projects, irrigation systems and water supply, water quality assessment, and navigation systems, among others (Blanco et al. 2013, Castellarin et al. 2013). The literature contains numerous examples of regional models, defined through multiple regression, for estimating FDCs in which the parameters of the curves are related to the physical and climatic characteristics of a river basin (Mimikou & Kaemaki 1985, Viola et al. 2011, Pessoa et al. 2011, Shu & Ouarda 2012, Costa et al. 2012, Mendicino & Senatore 2013, Waseem et al. 2015, Swain & Patra 2017, Silva et al. 2019).

Mimikou & Kaemaki (1985) developed a regionalization study of flow duration curves in western and northwestern Greece. Viola et al. (2011) developed a regional model to estimate FDCs in watersheds in Sicily, Italy. Regional regression equations were developed to obtain FDCs from the morphological characteristics of the basins. Mendicino & Senatore (2013) analyzed the performance of seven models of regional FDCs (two statistics and five parametric) for 19 calibrated basins in a region of southern Italy known as Calabria. For the definition of the regional models, they used multiple regression analysis.

Shu & Ouarda (2012) applied logarithmicregression-based logarithmic interpolation (RBLI), a method used to simulate FDCs at locations with no information, to 109 streamflow gauge stations in the province of Quebec, Canada. Waseem et al. (2015) predicted the streamflow percentiles of FDCs by combining three traditional methods (i.e., drainage area ratio, inverse distance weighted, and the regression method) in which the ensemble hydrological prediction performed better than the three individual traditional techniques in eight ungauged catchments in Pakistan.

Swain & Patra (2017) bring together a comparative assessment of streamflow estimation in ungauged catchments using regional FDCs. Four regionalization techniques, including area-index, inverse distance weighted (IDW), kriging, and stepwise regression, were applied to 32 catchments in India to estimate daily streamflow. The area-index method performed the worst, perhaps because it considers only drainage.

Regarding the estimation of FDCs, some works in the Amazon are highlighted, for example, in the studies developed by Pessoa et al. (2011) and Costa et al. (2012), in which the authors estimated FDCs in the hydrographic regions of the Calha Norte and Xingu in the state of Pará through multiple regression models. Silva et al. (2019) produced models of regionalization of FDCs, proposing the grouping of watersheds as a function of the drainage area without the use of cluster analysis techniques.

However, to achieve effective regionalization, it is necessary to define regions with hydrologically homogeneous behaviors, i.e., to have hydrological similarity between the physical and climatic characteristics of the region. One method used to obtain results when dividing a study area into homogeneous regions is cluster analysis, where the primary purpose is to aggregate objects based on a measure of the similarity of their characteristics. Among the methods of cluster analysis used in hydrology to obtain homogeneous regions, the hierarchical agglomerative method (Tsakiris et al. 2011, Rianna et al. 2011, Farsadnia et al. 2014, Awan 2015) and the Fuzzy C-Means (FCM) partitioning method (Sadri & Burn 2011, Satyanarayana & Srinivas 2011, Dikbas et al. 2012, Goyal & Gupta 2014, Senent-Aparicio et al. 2017, Beskow et al. 2016, Gomes et al. 2019) are highlighted.

Specifically, Sadri & Burn (2011) used the FCM procedure in the delimitation of homogeneous regions in the Canadian provinces of Alberta, Saskatchewan and Manitoba. The methodology was applied to the hydrological records of 36 streamflow monitoring sites based on bivariate criteria (severity and duration). The authors confirmed the importance of this methodology in helping to delimit homogeneous regions.

Satyanarayana & Srinivas (2011) presented an approach based on FCM cluster analysis in which it was possible to identify homogeneous regions of rainfall in India using large-scale atmospheric variables, as well as localization attributes and rainfall seasonality. Dikbas et al. (2012), in a study of rainfall series classification of 188 rainfall stations in Turkey, applied the FCM method and identified six hydrologically homogeneous regions based on the variables total annual precipitation, coefficient of variation of total annual precipitation, latitude and longitude.

Goyal & Gupta (2014) identified four homogeneous precipitation regions in northeastern India, using FCM. Senent-Aparicio et al. (2017) applied the Fuzzy C-Means and Fuzzy Minimals algorithms to assess the effectiveness of both algorithms in the identification of hydrologically homogeneous regions for flood frequency analysis of the watersheds in Alto Genil (southern Spain).

Beskow et al. (2016) evaluated the performance of artificial intelligence techniques (K-means. Partitioning Around Medoids. K-harmonic means, Fuzzy C-means and Genetic K-means) for the formation of hydrologically homogeneous regions in the State of Rio Grande do Sul (Brazil) for the regionalization of low flow - Q90. Gomes et al. (2019) identified three homogeneous precipitation regions in the hydrographic basin of Tocantins-Araguaia in Brazilian Amazonia using the Fuzzy C-means method and physical-climatic variables such as location (latitude and longitude), altitude, and precipitation.

In the context of the Amazon, there is a lack of work on the regionalization of flow continuity curves and the identification of homogeneous flow regions, with only studies by Pessoa et al. (2011), Costa et al. (2012) and Silva et al. (2019) in the state of Pará, Amazônia, Brazil. Therefore, the premise of this research is to minimize the problem of the region (absence of hydrological data), providing regional models that will allow us to simulate the permanence curve and not just the maximum, medium or low flow. According to the findings of Boscarello et al. (2016), it is important to define homogeneous regions, because the mean absolute percentage error decreased from 11% to 7% in the FDC estimates in 46 catchments in Italy.

The objective of the current study is to develop a methodology for the regionalization of FDCs in which the homogeneous regions are defined through the FCM method and regional models are defined through multiple regression. The main motivation for this work is the application of this methodology to the Amazon region, which, for the most part, is devoid of records of streamflow data, making it difficult to plan and manage the water resources of the world's largest river basin.

MATERIALS AND METHODS

Study area and data

The study area involves watersheds located in the Amazon between $5^{\circ}N-18^{\circ}S$ and $42^{\circ}W-74^{\circ}W$ and encompasses multiple states comprising the Brazilian Amazon (Acre, Amapá, Amazonas, Mato Grosso, Pará, Roraima, Rondônia, Tocantins and part of the state of Maranhão). In addition, the basins extend to neighboring countries (Venezuela, Colombia, Peru and Bolivia and French Guiana) (Figure 1).

Initially, 208 streamflow and 208 rainfall gauge stations were selected. These stations belong to the hydrometeorological network of the Hydrological Information System (HIDROWEB) of the Agência Nacional de Águas - ANA (2015) (http://hidroweb.ana.gov.br/). The



Figure 1. Amazonia and the spatial distribution of the streamflow and rainfall gauge stations used in the study.

stations were chosen after considering the distributions of existing data and historical series from 1975 to 2012. The data were used to calculate the mean annual precipitation P (mm), and the FDC for each station. The drainage area A (km^2) of the basins, the length of the main river L (km) and the head of the river H (m) were delimited by Geographic Information System (GIS) software using the digital elevation model (DEM) available at http://www.relevobr.cnpm.embrapa.br.

Fuzzy C-means (FCM) algorithm

The FCM algorithm is a multivariate data analysis technique that replaces the binary configuration of classical set theory with membership intervals in which an element belongs to one or more sets with a certain degree of pertinence between 0 and 1. This property is highly effective for the grouping of hydrological variables, as previously mentioned. The FCM algorithm was implemented in MATLAB 7.1 software through the "Fuzzy Logic Toolbox". This tool allows us to use a function known as fcm, which functions as an algorithm. As a stopping criterion, a minimum error of ε = 0.0001 and a maximum number of iterations tmax = 200 were used.

Data partitioning into fuzzy clusters is achieved by minimizing the objective function Jm (Equation 1), which then assists in verifying the convergence of the FCM algorithm. This function depends on the fuzzification parameter, m, which, with values between 1.25 and 2 (Ross 1995), guarantees effective performance. At each iteration t of the FCM algorithm, a rate of $J_m^{(t)}$ is calculated by means of J_m . $J_m^{(t)}$ is subtracted from $J_m^{(t-1)}$ to give ΔJ_m . If ΔJ_m is close to zero, it is an indicator that the algorithm is converging. In Equation 4, V = (v1, ..., c) is a vector containing the centroids of the clusters, in which $v_i(i = 1, ..., c) \in \Re^p$. Thus, each value x_k is evaluated according to its proximity to each centroid v_i . This comparison is made using the Euclidean distance between x_k and v_i . Specifically, u_{ik} is the degree of pertinence of x_k in cluster i. The centroids of the clusters are given by Equation 2, and the degrees of pertinence are given by Equation 3.

$$\nu_{i}^{(t)} = \sum_{k=1}^{n} \left(u_{ik}^{(t)} \right)^{m} x_{k} \swarrow \sum_{k=1}^{n} \left(u_{ik}^{(t)} \right)^{m} \qquad (2)$$

$$u_{jk}^{(t+1)} = 1/\sum_{j=1}^{C} \left(\frac{\|x_k - v_j^{(t)}\|^2}{\|x_k - v_j^{(t)}\|^2} \right)^{\frac{2}{m-1}}$$
(3)

The FCM algorithm may be summarized in the following steps (Farsadnia et al. 2014):

- Choose values for c (number of groups), m (fuzzification parameter) and stop criterion ε (error);
- 2. Randomly generate the array $U^{(0)}$ complying with the restrictions;
- 3. Assign the value o to the iteration counter;
- 4. Calculate the centroids (Equation 2), the objective function J_m (Equation 1) and the degrees of pertinence (Equation 3);
- 5. Compare the partition matrices $U^{(t)}$ and $U^{(t+1)}$. If $|U^{(t+1)} U^{(t)}| < \varepsilon$, finalize the algorithm; otherwise, return to step 3 by

incrementing the iteration counter t = 1, 2, n.

The clusters (homogeneous regions) were determined by the distribution of the variables: drainage area (km^2) , average long period streamflow (m^3/s) , mean annual precipitation (mm) and river length (km).

PBM INDEX

The PBM index was proposed by Pakhira et al. (2004). It is used to validate the clustering formed through the application of the FCM algorithm. This index is defined as the product of three factors (Equation 4). The maximization of this index ensures that the partition has the fewest possible clusters.

$$PBM(K) = \left(\frac{1}{K} \times \frac{E_1}{E_k} \times D_k\right)^2 \tag{4}$$

where K is the number of clusters. Additionally,

$$E_k = \sum_{k=1}^k E_k \tag{5}$$

such that

$$E_k = \sum_{j=1}^n u_{kj} \|X_j - Z_k\|$$
(6)

and

$$D_{k} = max_{i,j=1}^{k} \|Z_{i} - Z_{j}\|$$
(7)

where n is the total number of data values analysed, $U(X) = \left[u_{kj}\right]_{k \times n}$ is a partition array, and Z_k is the center of the kth cluster.

L-moments approach

Hosking & Wallis (1997) proposed a homogeneity test (H test) based on L-moment ratios (L-Cv, L-Cs, and L-Ck) for testing the homogeneity of the groups identified by cluster analysis. It is statistically evaluated by using a regional homogeneity test to determine if the groups determined with cluster analysis are homogeneous or not. Several studies in hydrology (Drissia et al. 2019, Gomes et al. 2019, Ghiaei et al. 2018, Dikbas et al. 2013) have used this test to confirm the homogeneity of a given region.

L moments are defined as linear combinations of probability weighted moments (PWM) of the time series (Hosking & Wallis 1997, Hosking & Wallis 1993). The first four L moments are as follows:

$$\lambda_1 = \beta_0 \tag{8}$$

$$\lambda_2 = 2\beta_1 - \beta_0 \tag{9}$$

$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 \tag{10}$$

$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \tag{11}$$

where β_0 , β_1 , β_2 and β_3 are the first four probability weighted moments (PWMs). L moment ratios (LMR) were calculated using Equations 12-14.

$$\tau_2 = \frac{\lambda_2}{\lambda_1} \tag{12}$$

$$\tau_3 = \frac{\lambda_3}{\lambda_2} \tag{13}$$

$$\tau_4 = \frac{\lambda_4}{\lambda_2} \tag{14}$$

where τ_2 is the L coefficient of variation (L-Cv), τ_3 is the L coefficient of skewness (L-Cs), τ_4 is the L coefficient of kurtosis (L-Ck). The τ_1 is considered as the average of the observed long period streamflow series.

According to Hosking & Wallis (1993, 1997), the test H statistic is used assess the homogeneity of a region based on L-moments and is a measure of heterogeneity (H_1 for L-Cv, H_2

for the combination of L-Cv and L-Cs, and H_3 for the combination of L-Ck and L-Cs) that compares sintersite variation in sample L-moments for a group of sites that would be expected in a homogeneous region. The heterogeneity measure (H_R) is defined by Equation 15.

$$H_{k} = \frac{(\nu_{\chi} - \mu_{\nu\chi})}{\sigma_{\nu\chi}} \tag{15}$$

where V_k is the weighted standard deviation of the L-Cv of the variation values, μ_{Vk} is the average of these values and ς_{Vk} is the standard deviation of the values obtained from the simulation.

According to the test of significance, which was proposed by Hosking & Wallis (1997), if H < 1, the region is considered "acceptably homogeneous", if $1 \le H < 2$, the region is "possibly homogeneous," and finally, if $H \ge 2$, the region should be classified as "definitely heterogeneous".

FDC fit

According to Castellarin et al. (2007), an FDC complements the empirical cumulative distribution function of daily streamflows based on the complete streamflow record available for the basin of interest.

To construct FDC, some authors (Viola et al. 2011, Ganora et al. 2009, Castellarin et al. 2007) recommend the use of a procedure that consists of two steps: (1) the observed streamflows q_i , i = 1, 2, ..., N, are sorted in descending order to produce a streamflow set q(1), i = 1, 2, ..., N, where N is the length of the sample, e $q_{(1)}$ and $q_{(N)}$ are the largest and the smallest observed streamflow events, respectively; (2) each ordered observation $q_{(i)}$ is plotted against its corresponding duration D_i , which is generally dimensionless and coincides with an estimate, p_i , of the probability of exceedance of $q_{(i)}$. In

the estimation of p_i , the Weibull plotting position (WPP) is used according to Equation 16.

$$P_i = P(Q > q_{(i)}) = \frac{i}{N+1}$$
 (16)

For a better graphical visualization of the fit of the models, 25 pairs - Q (m^3/s) x D (duration %) - were selected for each streamflow gauge station belonging to the hydrologically homogeneous regions previously defined. These 25 pairs were divided in intervals of 4% until reaching 100%; that is, 4, 8, 12 ... 100%.

Mathematical functions (Equations 17-22), used as models, were fitted to the observed FDC of the 208 streamflow gauge stations (Mimikou & Kaemaki 1985, Pessoa et al. 2011, Costa et al. 2012, Silva et al. 2019) as follows:

$$Linear \qquad Q = a - b.D \qquad (17)$$

Power
$$Q = a.D^{-b}$$
 (18)

Exponential
$$Q = a.e^{(-b.D)}$$
 (19)

Logarithmic Q = a - b.ln.D (20)

Quadratic
$$Q = a - b.D + c.D^2$$
 (21)

Cubic
$$Q = a - b.D + c.D^2 - d.D^3$$
 (22)

where Q (m^3/s) is the observed streamflow rate; D (%) is the equaled or exceeded duration; and a, b, c and d are the parameters resulting from the fit, which were calculated using the least squares method. In this case, the streamflow rate Q is the dependent variable and the duration D is the independent variable.

Performance criteria

To analyze the quality and performance of the model fits, we adopted the relative mean square error (Equation 23), the coefficient of determination (Equation 24), and the best-fit plot between the observed and simulated FDC.

$$\varepsilon = n^{-1} \left[\sum_{i=1}^{n} \left(\frac{y_i - \widehat{y_i}}{y_i} \right)^2 \right]^{\frac{1}{2}}$$
(23)

where y_i is the observed daily streamflow rate; $\hat{y_i}$ is the estimated value of the streamflow by the model; and n is the total number of observations.

$$R^{2} = \frac{\left(\left[\widehat{\beta}\right]^{T} \cdot \left[X\right]^{T} \cdot \left[Y\right] - n\overline{Y}^{2}\right)}{Y^{T} \cdot \left[Y\right] - n\overline{Y}^{2}}$$
(24)

where [Y] is a vector (n x 1) of the observations of the dependent variable, [X] is an array (n x P) with the n observations of each of the independent P variables, and [β] is a vector (P x 1) with unknown parameters. The coefficient of determination (R^2) describes the proportion of the variance of the measured data explained by the model. It ranges from 0 to 1, where values close to 1 indicate less error variation and values equal to or greater than 0.5 are considered acceptable (Santhi et al. 2011, Van Liew et al. 2007).

Regionalization

These models were constructed by means of multiple regression among the parameters (a, b, c and d) defined in the fit phase (Equations 17-22) in relation to the morphoclimatic characteristics of the river basins. These parameters explain the spatial variation in streamflow rates considering the drainage area A (km^2), the mean annual precipitation P (mm), the length L of the river (km) and the head of the river H (m). The regression equations applied were as follows:

$$V = \beta_0 + \beta_1 A + \beta_2 P + \beta_3 L + \beta_4 H$$
 (25)

$$V = \beta_0 . A^{\beta_1} . P^{\beta_2} . L^{\beta_3} . H^{\beta_4}$$
 (26)

$$V = \beta_0 \mathcal{A}^{\beta_1} \mathcal{P}^{\beta_2} \cdot \left(\frac{H}{L}\right)^{\beta_3}$$
(27)

$$V = \beta_0 P^{\beta_1} \left(\frac{A}{L}\right)^{\beta_2} H^{\beta_3}$$
(28)

where V is the dependent variable that represents the parameters of the FDC and β_0 , β_1 , β_2 , β_3 and β_4 are the regression coefficients determined by the least squares method.

The morphoclimatic characteristics, as well as the Equations 25-28, were chosen because they are the most commonly used in studies of the regionalization of FDCs in the Amazon (Pessoa et al. 2011, Costa et al. 2012) and are similar to the method of regionalization of FDC used by Mimikou & Kaemaki (1985) in watersheds of the western and northwestern regions of Greece. According to Duarte & Pessoa (2017), the variables drainage area, precipitation, river length and slope represented a frequency of use of 95.2%, 90.5%, 42.9% and 38.1%, respectively, in the regionalization studies of flow in the years 2010 to 2016.

The best model to be used for each hydrologically homogeneous region was evaluated by calculating the determination coefficient, R^2 (Equation 24). The F_{total} test (Equation 29) was also applied to verify the existence of a significant relationship between the dependent variable and the independent variables. The critical rate was obtained using the Snedecor's-F distribution.

$$F_{\text{total}} = \frac{\frac{[\widehat{\beta}]^{\mathsf{T}}[\mathsf{X}]^{\mathsf{T}}[\mathsf{Y}] - n\overline{\mathsf{Y}}^2}{\frac{P}{\frac{[\mathsf{Y}]^{\mathsf{T}}[\mathsf{Y}] - [\widehat{\beta}]^{\mathsf{T}}[\mathsf{X}]^{\mathsf{T}}[\mathsf{Y}]}{n-P-1}}$$
(29)

The model was considered to be statistically significant when the calculated rate of $F_{total} > F(\alpha, P, n - P - 1)$ at a significance level of 5%. ($\alpha = 0.05$).

Validation

The lack-knife cross validation method (Castellarin et al. 2007, 2009, Rianna et al. 2011) was used to validate the regionalization models. The procedure, summarized by Castellarin et al. (2007), consists of a repeated analysis of the results, excluding a station from the regression with the purpose of validating the model. The choice of this procedure is because the regionalization methodology to will repeated for all analyzed stations. Thus, each estimated curve may be compared to an observed curve to check the model errors. The Jack-knife technique was associated with the calculation of the NASH efficiency coefficient (Nash & Sutcliffe 1970) (Equation 30), which is described as follows:

$$NASH = 1 - \left[\frac{\sum_{i=1}^{n} (Y_{i}^{obs} - Y_{i}^{sim})^{2}}{\sum_{i=1}^{n} (Y_{i}^{obs} - \overline{Y}_{obs})^{2}}\right]$$
(30)

where Y_i^{obs} is the observed streamflow rate, Y_i^{sim} is the estimated streamflow rate and \overline{Y}_{obs} is the average of the observed streamflow rates. Rates between 0 and 1 are generally seen as acceptable levels of model performance, while rates below 0 indicate unacceptable performance (Nash & Sutcliffe 1970).

In addition to the NASH, we also used the relative mean square error (Equation 23) and the model evaluation statistic RSR (observations standard deviation ratio) (Equation 31). The RSR is calculated by dividing the square root of the sum of the quadratic errors by the standard deviation of the observations.

$$RSR = \left[\frac{\sqrt{\sum_{i=1}^{n} (Y_i^{obs} - Y_i^{sim})^2}}{\sqrt{\sum_{i=1}^{n} (Y_i^{obs} - \overline{Y}_{obs})^2}}\right]$$
(31)

The RSR varies from 0 to ∞ , where zero is the ideal value, indicating a perfect performance of the model in the simulation and zero residual variation or RMSE. Moriasi et al. (2007) recommend satisfactory streamflow-rate simulation hydrological models with a Nash-Sutcliffe (NASH) efficiency coefficient of greater than 0.50 and an RSR of less than 0.70. Figure 2 summarizes the methodology used in the present study.

RESULTS AND DISCUSSION

Homogeneous regions via FCM

Table I shows the PBM index values as a function of the cluster number (c) and the fuzzification parameter (m). Each column in Table I provides values corresponding to m ranging from 1.5 to 2.0. Figure 3 shows the graphs for these results. As seen in Table I and Figure 3, the PBM validation index had its value maximized for c = 10 and m = 1.6.

Pal & Bezdek (1995) noted that the FCM provides better performance for m (fuzzifier) in the range of 1.5–2.5. Gomes et al. (2019) observed better performance for the FCM algorithm with values of m = 1,9 and c = 3, which were validated through the PBM index. However, Srinivas et al. (2008) and Goyal et al. (2014) defined the best values of m = 1.5 and m = 1.7, respectively, but these values were validated by other methods.

The results demonstrate that the application dataset is best clustered into 10 groups. In this case, the algorithm reached the stop condition in 10 iterations (Figure 4). For the first iteration of the algorithm, the objective function jm provided the value of 6.79×10^{12} , and for the last iteration, the calculated value was equal to 4.44×10^{10} .

Figure 5 reflects an important contribution of the FCM algorithm since it can estimate the average streamflow rate of a river as a function of the homogeneous region, the drainage area and the average annual rainfall.

Figure 6 shows the spatial distribution of the hydrologically homogeneous streamflow regions found in the Amazon by means of the FCM algorithm. As seen by the clusters formed, geographic contiguity is not necessary to define a hydrologically homogeneous region (Rao & Srinivas 2006).

It is also noted that regions 6 to 8 grouped only a few stations, representing 6% of the total number of stations (Table II). Regions 1 to 4 grouped 90% of all of the stations. These regions, which show hydrological similarity, are distributed across the majority of the Amazonian territory, in comparison with the other regions. However, some regions of the Brazilian Amazon were not grouped due to the scarcity of data.

The FCM algorithm has been widely used in hydrology to identify homogeneous regions, as in studies by Beskow et al. (2016), who used artificial intelligence (AI) techniques with measures of the low seasonality of streamflow. The authors demonstrated the efficiency of the FCM algorithm in the identification of homogeneous regions, aiming to regionalize Q90 in southern Brazil. Gomes et al. (2019) identified three homogeneous rainfall regions for a watershed in the Amazon. Rao & Srinivas (2006) evaluated the FCM method for watersheds in Indiana, USA, and found great potential of that algorithm to determine homogeneous regions for modeling annual maximum stream flows. Sadri & Burn (2011) also obtained promising results from the FCM technique to delineate regions for low stream flow regionalization in watersheds located in Canada.

Regional homogeneity test based on L-moments (H test)

The hydrological homogeneity of the clusters identified by the FCM algorithm should be tested for use in streamflow estimation studies. A regional homogeneity test based on the L-moments proposed by Hosking & Wallis (1993, 1997) was used in this study. H (H1, H2 and H3) values were calculated for clusters 1-8



Figure 2. Summary of the methodology.

Table I. Results o	f the PBM index	to the clustering	of the FCM algorithm.

Number of cluster	PMB-index							
Number of cluster	m=1.5	m=1.6	m=1.7	m=1.8	m=1.9	m=2.0		
2	4.65E+09	4.64E+09	1.02E+12	1.01E+12	1.00E+12	4.64E+09		
3	2.38E+09	2.82E+10	2.37E+09	2.38E+09	2.55E+10	1.56E+11		
4	1.29E+11	1.76E+10	1.48E+09	4.59E+10	4.00E+11	1.61E+11		
5	1.01E+10	3.35E+10	1.47E+10	1.10E+12	3.08E+11	6.82E+10		
6	2.40E+10	1.99E+10	1.30E+10	9.37E+08	2.34E+10	9.54E+09		
7	1.80E+08	8.71E+08	2.12E+10	6.11E+09	1.51E+11	5.65E+09		
8	2.40E+10	6.76E+09	1.55E+10	2.90E+10	6.50E+08	5.54E+09		
9	3.28E+09	5.80E+09	5.87E+08	5.49E+10	6.51E+09	7.53E+09		
10	5.16E+09	1.32E+12	2.80E+09	4.99E+10	2.48E+09	1.21E+10		
11	1.48E+10	1.46E+09	1.77E+09	2.13E+10	5.61E+08	3.79E+10		
12	1.44E+10	3.52E+10	1.66E+10	1.01E+09	9.19E+09	1.20E+07		
13	1.21E+09	1.64E+10	3.35E+09	1.52E+10	7.62E+09	5.01E+08		
14	1.25E+09	5.03E+09	1.45E+09	4.13E+09	9.72E+08	3.27E+10		
15	5.64E+09	4.96E+09	1.73E+09	1.26E+09	1.42E+10	8.83E+09		



Figure 3. Results of the PBM index to the clustering of the FCM algorithm.



Figure 4. Convergence of the objective function for the 10 clusters.

An Acad Bras Cienc (2021) 93(1) e20190747 11 | 21



Figure 5. Clusters according to the drainage area, mean annual precipitation and average long-term streamflow.

Clusters	NS	%	Drainage Area (km²)	Average Streamflow (m³/s)	Average Rainfall (mm)
1	101	48.56	491-17,990	150	1837
2	55	26.44	18,394 - 47,038	925	1883
3	21	10.10	51,147 - 112,186	1824	2027
4	10	4.81	133,571 - 193,372	4646	2343
5	6	2.88	225,424 - 293,084	13014	2178
6	4	1.92	317,967 - 367,791	22882	2342
7	3	1.44	456,347 - 508,733	9083	1829
8	6	2.88	889,201 - 1,082,709	17661	2162
9	1	0.48	1,402,097	101158	2250
10	1	0.48	3,911,283	170013	1778
Total	208	100			

Table II. Clustering x Data distribution.

NS – number of stations.

determined by the FCM algorithm and are given in Table III. This test was not applied to groups 9 and 10, since each one was defined containing only one station, though these stations represent the basins with the largest drainage area present in the Amazon region. Based on the values of the H statistic (Table III), groups 1-8 were evaluated as "acceptably homogeneous", since all H values found were less than 1. Hence, regions 1-8 can be considered hydrologically homogeneous for all average long-period streamflows.

Ghiaei et al. (2008) applied the L-moments technique at all stages of regional rainfall



Figure 6. Spatial distribution of the hydrologically homogeneous streamflow regions found in the Amazon.

Region	The number of stations		H test		Homogonoity commont	
		H1	H2	H3	nomogeneity comment	
1	101	-0.1431	-0.1397	-0.00038	Acceptably homogeneous	
2	55	-0.1237	-0.1155	-0.00991	Acceptably homogeneous	
3	21	-0.0930	-0.1052	0.01407	Acceptably homogeneous	
4	10	-0.0874	-0.0869	0.00472	Acceptably homogeneous	
5	6	-0.0857	-0.0825	0.00152	Acceptably homogeneous	
6	4	-0.0651	-0.1197	-0.0005	Acceptably homogeneous	
7	3	-0.1066	-0.1053	0.00026	Acceptably homogeneous	
8	6	-0.0948	-0.0962	-0.00694	Acceptably homogeneous	
9	1	-	-	-	has not been applied	
10	1	-	-	-	has not been applied	

Table III. The results of the regional homogeneity test for clusters defined by the FCM algorithm.

analysis in Turkey, including the determination of homogeneous regions, in addition to fitting and estimating parameters from appropriate distribution functions in each homogeneous region. Likewise, Gomes et al. (2019) confirmed the homogeneity of three regions of precipitation in the Hydrographic Region of Tocantins-Araguaia using the heterogeneity H test.

FDC fit

No regional models were developed for homogeneous regions 5 - 10 (Figure 6) due to the small number of existing streamflow gauge stations. Figure 7 presents examples of the fit of the mathematical model that best estimated the FDC for the homogeneous regions 1 – 4.

Table IV shows the mean of the coefficients of determination (R^2) and the relative mean squared errors (ε %) for the 4 homogeneous regions of all the models used in the FDC fit. The fits were found to be satisfactory for the evaluation of the means of R^2 and ε % and for the graphical analysis of the fit of the best model to the observed FDC (Figure 7). The Table III and the Figure 7 shows that for homogeneous region 1, the cubic model provided the best fit of the permanence curves. In regions 2, 3 and 4, the exponential model provided the best fit. It is important to note that, as seen in Figure 3, the best model of each homogeneous region was adjusted to the FDC approximately equally during periods of floods as well as during periods of droughts. This result indicates that the parameters a, b, c and d, since D (%) is known, satisfactorily explain the spatial variation in the streamflows based on the morphoclimatic characteristics of the basins of each homogeneous region.

These results are in agreement with those found by Silva et al. (2019), Costa et al. (2012) and Pessoa et al. (2011), which demonstrated the efficiency of the cubic and exponential models in the adjustment with the curves of permanence in the watersheds of the Amazon. In contrast, studies by Otache et al. (2016) and Shu & Ouarda (2012) obtained the best results for the exponential and logarithmic models. According

Models	Region 1		Region 2		Region 3		Region 4	
Models	R2	ε (%)						
Linear	0.92	8.19	0.87	9.08	0.89	7.67	0.91	6.96
Potential	0.90	4.52	0.85	6.54	0.84	6.17	0.79	7.31
Exponential	0.95	2.28	0.97	2.17	0.98	1.31	0.98	2.11
Logarithmic	0.97	3.58	0.96	4.66	0.94	4.12	0.93	5.09
Quadratic	0.95	4.20	0.97	3.87	0.97	2.66	0.98	2.97
Cubic	0.98	1.51	0.97	2.98	0.98	1.74	0.98	2.70

Table IV. Mean of R2 and ε % of each model in the FDC fit for the 4 homogeneous regions.



2 . . .

Figure 7. Fit of the FDC observed for Regions 1-4.

to Silva et. al. (2019), this difference can be explained, since the adoption of a single model for all situations may not be adequate, since the physical and climatic characteristics of each basin are unique.

Regional models

Table V shows the best regional models for estimating the parameters a, b, c and d and, consequently, the best models to estimate the streamflow continuity curves for the 4 homogeneous regions (HR).

The Table V also shows the performance criteria for the establishment of the models, i.e., R^2 and F_{total} . Thus, in Region 1, the cubic model was the best. In Region 2, the exponential model was the best. These models presented R^2 values above 0.50, in addition to a significant relationship between the parameters of the models and the explanatory variables to 5% of significance. This result is explained by the rates found in the Snedecor's F distribution. which are equal to 2.70 and 2.79, respectively, which are smaller than the F_{total} found for Equations 32-35 and Equations 36-37. In regions 3 and 4, which have 29 and 10 streamflow gauge stations, respectively, the Snedecor's F distribution rates are 3.20 and 4.76 at a significance of 5%. It is observed in Table IV that Equations 38-41 of the regional models suggest that these regions did not pass the F_{total} test and presented R^2 values for parameter b below 0.50. However, the results should not completely rule out a regression equation because regression coefficients can show significant correlation. However, the regional models of these data revealed unsatisfactory results, probably due to the application of multiple regressions in very small clusters. Thus, validation was performed for the regionalized FDC models, including the unsatisfactory results.

Validation of regional models

Figures 8, 9, 10 and 11 presents a comparison of the rates of the average relative mean square error (ε %), the Nash-Sutcliffe efficiency coefficient (NASH) and the ratio by standard deviation (RSR). The figures also shows the mean values of NASH and ε % for each streamflow gauge station from the homogeneous regions 1, 2, 3 and 4 that was removed from the regression according to the Jack-knife method.

An analysis of the data in Figure 8 verified that only one streamflow gauge station presented an ε % greater than 20%, with a mean ε % of 7.19. It can be observed that only seven stations (E3, E21, E154, E175, E185 and E194 - 8% of the sample) can be considered unacceptable because they exhibited NASH values of less than o. With the values of NASH and RSR for each station removed, it was determined that 81.61% of the streamflow gauge stations have NASH coefficients higher than 0.50 and RSR lower than 0.70. These results demonstrate that in more than 80% of the cases, there was a satisfactory fit of the data observed and that simulated by the model. In less than 20% of the cases, the NASH values were less than 0.50, and the RSR was greater than 0.70. The performance of the model can be accepted as good considering these rates of NASH and RSR (Moriasi et al. 2007).

Similar results were found in the studies conducted by Swain & Patra (2017), in which an average Nash-Sutcliffe value of 0.60 indicated better regional model performance when comparing four techniques; in the study, regionalization flow duration curves (area index, inverse distance weighting (IDW), kriging and stepwise regression methods) were applied to assess 32 catchments in India. Silva et al. (2019) formed 3 homogeneous regions as a function of the drainage area without the use of cluster analysis techniques. Regional models presented

HR	Regionalization models	Eq.	R2	Ftotal
1	$a = 1.30 \times A^{0.689} P^{-0.060} \left(\frac{H}{L} \right)^{-0.133}$	(32)	0.81	108.64
	$b = 1253.8 \times A^{0.604} \times P^{-0.756} \times \left(\frac{H}{L}\right)^{-0.183}$	(33)	0.62	45.22
	$c = 1619.1 \times A^{0.673} \times P^{-0.816} \times \left(\frac{H}{L}\right)^{-0.064}$	(34)	0.53	31.56
	$d = 351.1 \times A^{0.683} \times P^{-0.719} \times \left(\frac{H}{L}\right)^{-0.024}$	(35)	0.51	26.87
2	$a = -2175.8 + 0.059 \times A + 1.166 \times P + 40.44 \times \left(\frac{H}{L}\right)$	(36)	0.61	9.44
	$b = 1.96 - 3 \times 10^{-6} \times A + 0.0002 \times P - 0.024 \times \left(\frac{H}{L}\right)$	(37)	0.54	2.92
3	$a = -5384.6 + 0.057 \times A + 2.71 \times P + 4154.1 \times \left(\frac{H}{L}\right)$	(38)	0.52	2.05
	$b = 2.54 - 1.3 \times 10^{-5} \times A + 5.8 \times 10^{-5} \times P + 6.73 \times \left(\frac{H}{L}\right)$	(39)	0.31	0.53
4	$a = -38707.6 + 0.087 \times A + 13.37 \times P + 492236.4 \times \left(\frac{H}{L}\right)$	(40)	0.64	1.84
	$b = 2.94 - 1.0 \times 10^{-5} \times A + 0.0004 \times P + 24.61 \times \begin{pmatrix} H \\ T \end{pmatrix}$	(41)	0.24	0.79

Table V. Models of regionalization by homogeneous region.



Figure 8. Graphical comparison of ɛ%, NASH and RSR for the homogeneous region 1.



Figure 9. Graphical comparison of ϵ %, NASH and RSR for the homogeneous region 2.

An Acad Bras Cienc (2021) 93(1) e20190747 17 | 21

Figure 10. Graphical comparison of ε %, NASH and RSR for the homogeneous region 3.

Figure 11. Graphical comparison of £%, NASH and RSR for the homogeneous region 4.

satisfactory performance in the estimation of FDC because the average values of the Nash-Sutcliffe coefficient of both models were higher than 0.60 and the relative mean square error was less than 20%.

In research conducted by Boscarello et al. (2016), Mendicino & Senatore (2013) and Castellarin et al. (2007), the cross-validation method was used to evaluate the performance of the regionalization of the flow duration curves model developed through multiple linear regression. Nash- Sutcliffe values lower than 0.50 were considered a weak fit between the simulated and observed curves, so they were not satisfactory. Figures 9-11 show the graphs that compare the criteria used for the validation of the efficiency of the regional models for the homogeneous regions 2-4.

An analysis of Figures 9-11 indicates unacceptable performance in two cases for regions 3 (E63 and E74) and 4 (E102 and E117) and one in region 2 (E24) because the NASH coefficients were less than 0. The averages of the NASH coefficients for the three regions were higher than 0.60, and the averages of ε % were close to 10%, at 8.74%, 9.93% and 12.26%. This result is considered a satisfactory performance for the regional models. It can also be observed that the regional models identified for the three regions presented satisfactory performance because the NASH coefficient values were higher than 0.50 and RSR values were less than 0.70.

CONCLUSION

The methodology developed exhibited satisfactory performance, taking into account the analysis of the results of the criteria analyzed. The homogeneous regions were well defined by the FCM method, and the regional models, through multiple regressions, were able to satisfactorily simulate the FDCs. which can be confirmed by the good graphical representation of these curves as a function of the data observed. This satisfactory result is important in a region that, for the most part, has insufficient streamflow data, making it difficult to plan and manage water resources. However, it is observed that this same lack of streamflow data was related to the main limitation of the study. That is, it is difficult to apply regional models due to data scarcity. Even with this limitation in quantity, our methodology can be applied to 4 homogeneous regions (Regions 1-4), representing 89.91% of the total streamflow gauge stations considered in this study. Therefore, through the regionalization models developed and with the input of available data, FDCs can be estimated in the majority of the Amazon region (approximately 75%). Therefore, the methodology presented here can be a valuable tool to support issues involving the planning and management of water resources, which depend on the knowledge of FDCs. However, it was not possible to apply the methodology to the streamflow gauging stations in homogeneous regions (Regions 5-10) representing 25% of the overall region. In this case, other methodologies should be developed for the estimation of FDCs.

Acknowledgments

The authors would like to thank Agência Nacional de Águas (ANA) for kindly providing rainfall and streamflow data for the current analysis. The first author would like to thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPg) for the PhD degree scholarship. The second author would like to thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for funding the research productivity grant (Process 304936/2015-4). The third author would like to thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the Master's degree scholarship. The authors would like to thank the Pró Reitoria de Pesquisa e Pós-Graduação (PROPESP) and the Fundação de Amparo e Desenvolvimento da Pesquisa (FADESP) of the Universidade Federal do Pará for their support through grant number PAPQ 2018.

REFERENCES

AGÊNCIA NACIONAL DE ÁGUAS - ANA. 2015. Acessado em: fevereiro de 2015. Acessível em: http://www.ana.gov.br.

AWAN JA, BAE D & KIM K. 2015. Identification and trend analysis of homogeneous rainfall zones over the East Asia monsoon region. Int J Climatol 35(7): 1422-1433.

BESKOW S, MELLO CR, VARGAS MM, CORRÊA LL, CALDEIRA TL, DURÃES MF & AGUIAR MS. 2016. Artificial intelligence techniques coupled with seasonality measures for hydrological regionalization of Q90 under Brazilian conditions. J Hydrol 541: 1406-1419.

BLANCO CJC, SANTOS SSM, QUINTAS MC, VINAGRE MVA & MESQUITA ALA. 2013. Contribution to hydro modelling of small Amazonian catchments: application of rainfall-runoff models to simulate flow duration curves. Hydrol Sci J 58(7): 1423-1433.

BOSCARELLO L, RAVAZZANI G, CISLAGHI A & MANCINI M. 2016. Regionalization of flow-duration curves through catchmen classification with streamflow signatures and physiographicclimate indices. J Hydrol Eng 21(3): 05015027.

CASTELLARIN A, CAMORANI G & BRATH A. 2007. Predicting annual and long-term flow duration curves in ungauged basins. Adv Water Resour 30(4): 937-953.

CASTELLARIN A ET AL. 2013. Prediction of flow du- ration curves in ungauged basins. In: BLÖSCHL G, SIVAPALAN M, WAGENER T, VIGLIONE A & SAVENIJE H (Eds). Runoff Prediction in Ungauged Basins: Synthesis Across Processes, Places and Scales. Cambridge University Press. 135-161.

CASTIGLIONI S, CASTELLARIN A, MONTANARI A, SKOIEN JO, LAAHA G & BLOSCHL G. 2009. Smooth regional estimation of

low-flow indices: physiographical space based interpolation and top-kriging. Hydrol Earth Syst Sci 15(3): 715-727.

COSTA AS, CARIELLO BL, BLANCO CJC & PESSOA FCL. 2012. Regionalização de curvas de permanência de vazão de regiões hidrográficas do estado do Pará. Rev Bras Meteorol 27(4): 413-422.

DIKBAS F, FIRAT M, CEM KOC A & GUNGOR M. 2012. Classification of precipitation series using fuzzy cluster method. Int J Climatol 32(10): 1596-1603.

DIKBAS F, FIRAT M, CEM KOC A & GUNGOR M. 2013. Defining homogeneous regions for streamflow processes in Turkey using a k-means clustering method. Arab journal Sci Eng 38(6): 1313-1319.

DRISSIA TK, JOTHIPRAKASH V & ANITHA AB. 2019. Flood Frequency Analysis Using L Moments: a Comparison between At-Site and Regional Approach. Water Resour Manag 33(3): 1013-1037.

DUARTE JM & PESSOA FCL. 2017. Estudo das características físico-climáticas mais significativas para a regionalização de vazões. XXII Simpósio Brasileiro de Recursos Hídricos – Florianópolis – SC.

FARSADNIA F, KAMROOD MR, NIA AM, MODARRES R, BRAY MT, HAN D & SADATINEJAD J. 2014. Identification of homogeneours regions for regionalization of watersheds by two-level self-organizing feature maps. J Hydrol 509: 387-397.

GANORA D, CLAPS P, LAIO F & VIGLIONE A. 2009. An approach to estimate nonparametric flow duration curves in ungauged basins. Water Resour Res 45(10): 1-10.

GHIAEI F, KANKAL M, ANILAN T & YUKSEK O. 2018. Regional intensity-duration-frequency analysis in the Eastern Black Sea Basin, Turkey, by using L-moments and regression analysis. Theor Appl Climatol 131(1-2): 245-257.

GOMES EP, BLANCO CJC & PESSOA FCL. 2019. Identification of homogeneous precipitation regions via fuzzy c-means in the hydrographic region of Tocantins-Araguaia of Brazilian Amazonia. Appll Water Sci 9(6): 1-12.

GOYAL MK & GUPTA V. 2014. Identification of homogeneous rainfall regimes in Northeast Region of India using fuzzy cluster analysis. Water Resour Manag 28: 4491-4511.

HOSKING JRM & WALLIS JR. 1997. Regional frequency analysis: an approach based on L-moments. Cambridge University Press Cambridge, 224.

HOSKING JRM & WALLIS JR. 1993. Some statistics useful in regional frequency analysis. Water Resour Res 29(2): 271-281.

LI M, SHAO Q, ZHANG L & CHIEW FHS. 2010. A new regionalization approach and its application to predict flow duration curve in ungauged basins. J Hydrol 389(1-2): 137-145.

MENDICINO G & SENATORE A. 2013. Evaluation of parametric and statistical approaches the regionalization of flow duration curves in intermittent regimes. J Hydrol 480: 19-32.

MIMIKOU M & KAEMAKI S. 1985. Regionalization of flow duration characteristics. J Hydrol 82(1-2): 77-91.

MORIASI DN, ARNOLD JG, VAN LIEW MW, BINGNER RL, HARMEL RD & VEITH TL. 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. Am Soc Agric and Biol Eng 50(3): 885-900.

NASH JE & SUTCLIFFE JV. 1970. River flow forecasting though conceptual models part I – A discussion of principles. J Hydrol 10(3): 282-290.

OTACHE MY, TYABO MA, ANIMASHAUN IM & EZEKIEL LP. 2016. Application of Parametric-Based Framework for Regionalisation of Flow Duration Curves. J Geo and Environ Prot 4 (5): 89-99.

PAKHIRA MK, BANDYOPADHYAY S & MAULIK U. 2004. Validity index for crisp and fuzzy clusters. Pattern Recognit 37(3): 487-501.

PAL NR & BEZDEK JC. 1995. On cluster validity for the fuzzy c-means model. IEEE Trasactions on Fuzzy Systems 3(3): 370-379.

PESSOA FCL, BLANCO CJC & MARTINS JR. 2011. Regionalização de curvas de permanência de vazões da região hidrográfica da Calha Norte no estado do Pará. Rev Bras Rec Hidr 16(2): 65-74.

RAO AR & SRINIVAS VV. 2006. Regionalization of watersheds by hybrid-cluster analysis. J Hydrol 318(1-4): 37-56.

RIANNA M, RUSSO F & NAPOLITANO F. 2011. Stochastic index model for intermittent regimes: from preliminary analysis to regionalisation. Nat Hazards Earth Syst Sci 11: 1189-1203.

ROSS TJ. 1995. Fuzzy Logic with Engineering Applications. McGraw-Hill, New York.

SADRI S & BURN DH. 2011. A fuzzy c-means approach for regionalization using a bivariate homogeneity and discordancy approach. J Hydrol 401(3-4): 231-239.

SANTHI C, ARNOLD JG, WILLIAMS JR, DUGAS WA, SRINIVASAN R & HAUCK LM. 2001. Validation of the SWAT model on a large river basin with point and nonpoint sources. J. Am. Water Resour Assoc 37(5): 1169-1188.

SATYANARAYANA P & SRINIVAS VV. 2011. Regionalization of precipitation in data sparse areas using large scale atmospheric variables – a fuzzy clustering approach. J Hydrol 405(3-4): 462-473.

SENENT-APARICIO J, SOTO J, PÉREZ-SÁNCHEZ J & GARRIDO J. 2017. A novel fuzzy clustering approach to regionalise watersheds with an automatic determination of optimal number of clusters. J Hydrol Hydromech 65(4): 359-365. SHU C & OUARDA TBMJ. 2012. Improved methods for daily streamflow estimates at ungauged sites. Water Resour Res 48 (2): 1-15.

SILVA RS, BLANCO CJC & PESSOA FCL. 2019. Alternative for regionalization of flow duration curves. J Appl Water Eng Res 7(3): 198-206.

SRINIVAS, VV, TRIPATHI S, RAO AR & GOVINDARAJU RS. 2008. Regional flood frequency analysis by combining self-organizing feature map and fuzzy clustering. J Hydrol 348(1-2): 148-166.

SWAIN JB & PATRA KC. 2017. Streamflow estimation in ungauged catchments using regional flow duration curve: comparative study. J Hydrol Eng 22(7): 04017010.

TSAKIRIS G, NALBANTINS L & CAVADIAS G. 2011. Regionalization of low flows based on canonical correlation analysis. Adv Water Resour 34(7): 865-872.

VAN LIEW MW, VEITH TL, BOSCH DD & ARNOLD JG. 2007. Suitability os SWAT for the conservation effects assessment project: A comparison on USDA-ARS experimental watersheds. J Hydrol Eng 12(2): 173-189.

VIOLA F, NOTO LV, CANNAROZZO M & LA LOGGIA G. 2011. Regional flow duration curves for ungauged sites in Sicily. Hydrol Earth Syst Sci 15(1): 323-331.

WASEEM M, AJMAL M & KIM T. 2015. Ensemble hydrological prediction of streamflow percentile at ungauged basins in Pakistan. J Hydrol 525: 130-137.

How to cite

PESSOA FCL, BLANCO CJC & GOMES EV. 2021. Regionalization of flow duration curves in the Amazon with the definition of homogeneous regions via fuzzy C-means. An Acad Bras Cienc 93: e20190747. DOI 10.1590/0001-3765202120190747.

Manuscript received on July 1, 2019; accepted for publication on September 10, 2019

FRANCISCO C.L. PESSOA¹

https://orcid.org/0000-0002-6496-9043

CLAUDIO J.C. BLANCO²

https://orcid.org/0000-0001-8022-2647

EVANICE P. GOMES³

https://orcid.org/0000-0001-9703-4837

¹Programa de Pós-Graduação em Engenharia de Recursos Naturais da Amazônia, Universidade Federal do Pará, PRODERNA/ITEC/UFPA, Rua Augusto Corrêa, 01, Guamá, 66075-110 Belém, PA, Brazil

²Universidade Federal do Pará, FAESA/ITEC/UFPA, Faculdade de Engenharia Sanitária e Ambiental, Rua Augusto Corrêa, 01, Guamá, 66075-110 Belém, PA, Brazil

³Programa de Pós-Graduação em Engenharia Civil, PPGEC/ITEC/UFPA, Rua Augusto Corrêa, 01, Guamá, 66075-110 Belém, PA, Brazil

Correspondence to: Claudio José Cavalcante Blanco

E-mail: blanco@ufpa.br

Authors contributions

Francisco Carlos Lira Pessoa contributed to the initial research study; in the acquisition, analysis and interpretation of data; the preparation of the manuscript; in the review of results and in the critical review. Claudio José Cavalcante Blanco contributed to the Study design and research; as well as contributions to the analysis of results and conclusions; in the correction and critical review. Evanice Pinheiro Gomes participated in the study design and initial research; in the review of results and critical review.

