JEE Journal of Ecological Engineering

Journal of Ecological Engineering 2023, 24(4), 357–367
https://doi.org/10.12911/22998993/161078
ISSN 2299–8993, License CC-BY 4.0

Received: 2023.01.16
Accepted: 2023.02.09
Published: 2023.03.01

Identification of Homogeneous Regions of Specific Minimum Flows in the State of Goiás, Brazil

Raviel Basso[1,2], Kássia Santana[1], Michelle Honório[1,3], Isabella Costa[1,3], Sanderson Leitão[4,5], Antonio Albuquerque[4,5,*], Paulo Scalize[1,2,3,5]

[1] School of Civil and Environmental Engineering, Federal University of Goiás, Av. Universitária, Quadra 86, Lote Área 1488 - Setor Leste Universitário, Goiânia 74605-220, Goiás, Brazil

[2] PPGEAS, School of Civil and Environmental Engineering, Federal University of Goiás, 74000-000 Goiânia, Goiás, Brazil.

[3] CIAMB, Federal University of Goiás, Avenida Esperança s/n, Câmpus Samambaia - Prédio da Reitoria, Goiânia 74690-900, Goiás, Brazil.

[4] Department of Civil Engineering and Architecture, GeoBioTec, University of Beira Interior, Calcada Fonte do Lameiro 6, 6200-358 Covilhã, Portugal

[5] FibEnTech, GeoBioTec-UBI, University of Beira Interior, Calcada Fonte do Lameiro 6, 6200-358 Covilhã, Portugal

* Corresponding author's e-mail: antonio.albuquerque@ubi.pt

ABSTRACT:
Hydrological information is essential for adequate water resources management as well as for water supply, energy supply, water allocation, among other services. However, this information does not always exist in quantity and quality to be used in hydrological or water management studies, and alternative methods are required to estimate minimum flows. Estimation based on homogeneous regions enables to transfer observation data from a known location to a location without data, but in the same region. Since the fluviometric stations in the state of Goiás (Brazil) are not uniformly distributed, the present work aimed at delimiting homogeneous regions of minimum flows, using the cluster grouping method with the K-means algorithm.Thus, 71 fluviometric stations with at least 5 years of continuous data were selected, obtained from the HIDROWEB system. In addition to the observed data, other variables were considered, such as drainage area, perimeter, specific minimum flows Q7,10, Q90, Q95 and average slope. The use of all these variables together with the observed data made it possible to determine,with great accuracy, 5 homogeneous regions of minimum flows based on the cluster analysis, enabling to obtain the minimum flows of reference for each region.In the selected homogeneous regions, it was possible to observe that the regions with the highest values of average slope presented smaller minimum flows, and the same could be observed under inverse conditions, i.e., lower values of average slope had higher minimum flows.It is also noteworthy that river monitoring is deficient in the center-south and center-north parts of the state of Goiás, making water resources management difficult. This fact indicates, therefore, the need to expand the river monitoring system throughout the state, especially in its southern and northern regions.

Keywords: Reference flows, permanence curve, K-means method, flow regionalization, water management.

INTRODUCTION

Water is fundamental to maintain life on the planet. Besides, water in sufficient quantity and with adequate quality is also a requisite for human activities in its multiple uses: human and animal consumption, health, agriculture, industry, energy generation, transportation, leisure and recreation, sewage, and effluent dilution, among others. Therefore, correct water resources management becomes crucial to the development of the countries and regions as well to conservation and maintenance of water sustainable for present and future generations.In this perspective, the availability of water resources in quantity and quality is fundamental for the occurrenceof economic activities, as well as for water supply for both human and animal consumption. Watersheds function as planning and management units, assuming a fundamental role when it comes to maintaining water resources sustainablefor present and future generations. The rational and integrated use of water for the prevention and defense against critical hydrological events (in the context of minimum flows), represent the aspects that are objectives of the main legal water resources management instrument in Brazil, the Water Resources National Policy, established by Federal Law No. 9.433/1997 (Brazil, 1997).

It is thus necessary to have suitable legal instruments to properly manage water resources and avoid water scarcity situations that may lead to water crisis, a condition that is more and more frequent in many parts of Brazil and the world. In this regard, the knowledge on the hydrological behavior of the place ~~of~~ interest is essential for water resources management and planning out to be carried out, since it plays an important role in several types of water useprojects, such as water supply, energy supply, among others (Cupak et al., 2017). However, the areas of interest do not always have enough information, for several reasons, thus, flow estimation methods are needed (Kim et al., 2016).According to Araújo & Rocha (2010),water resources assessment, in respect to water availability, can be performed through minimum, average and maximum flows, together with the variation of precipitation. Smakthin (2001) relates that minimum flows are responsible for indicating natural water availability in a watershed and are essential variables for water use management, for example, and may also influence the management of conflicts in water scarcity scenarios (Granemann et al., 2018).

It is noteworthy that for the state of Goiás, the State Water Resources Council (CERHi) resolution nº 22/2019 (Goiás, 2019), which replaced the old resolution nº 09/2005 (Goiás, 2005), defines as a reference, the minimum flow rate with guarantee of 95% in time (Q95). The same minimum reference flow has been adopted by several Brazilian states: Rio de Janeiro, EspíritoSanto, Paraná, Mato Grosso, Mato Grosso do Sul, among others. However, in the states of Minas Gerais and São Paulo, the reference flow used for management of water (water allocations) is the minimum flow with seven days of duration and ten years of return (Q7.10), while some Brazilian states adopt a flow rate that equaled or exceeded 90% of the time (Q90).On the other hand,the state of Goiás has an insufficient fluviometric monitoring network according to the World Meteorological Organization (WMO; 2008), for the adequate characterization of the behavior of the minimum flows of its water courses as it was noted by Basso et al (2022). In addition, in the existing stations, some problems are observed in the collected data, such as the presence of faults (missing data), inconsistent data and short periods, thus, the collected information is not sufficient to carry out accurate analyses. According to (Beskow et al., 2013), this fact occurs more frequently in watersheds of medium and small dimensions.

For this reason, one of the available alternatives is the estimation of flow through regionalization, a method used as a way of filling the deficit of hydrological data in places with little or no information (Eslamian and Biabanak, 2008; Samuel et al., 2011; Pruski, et al., 2012; Basso et al., 2022). The use of the method has shown satisfactory results and can provide flexibility in the decision-making process, but it does not replace the need to monitor hydrological variables (Maciel et al., 2019). Thus, to improve minimum flow regionalization methods, several studies seek to group hydrological variables into homogeneous regions, considered when there is evidence that the information can be transferred from one location to another (Zhang and Stadnyk, 2020). Delimitation of these regions can be performed in different ways, since there is no consensus on a single regionalization methodology to be applied in all situations (Hosking and Wallis, 1997; Razavi and Coulibaly, 2013).

Several national and international studies present different methodologies for the determination of homogeneous regions, such as frequency distribution analysis (Burn, 1989; Tucci et al., 1995), cluster analysis via the Ward hierarchical method (Nathan and McMahon, 1990; Kazemzadeh and Malekian, 2018), from fuzzy c-means (Hall and Minns, 1999; Gomes et al., 2018; Silva, 2018), hybrid methods (Gaál et al. 2009; Farsadnia et al., 2014; Nadoushani et al., 2018), the k-means method (Demirel et al., 2007; Kahya et al., 2007; Dikbas et al., 2013; Wu et al., 2015; Lücke and Forster, 2019), among others.

The k-means method has been widely used because it has advantages over other methods, such as having low complexity, fast computation, having the capacity to process large data sets and the adjustable grouping association (Aytaç, 2020). According to Dikbas et al. (2013) the method can be successfully applied in the classification of maximum annual flows and in the identification of hydrologically homogeneous regions. On the basis of homogeneous regions, it is possible to transfer data from one location to another, where these do not exist or are scarce, facilitating management in watersheds without monitoring (Nadoushani et al., 2018). Therefore, the present paper aimed to fill the gaps in different watershed with low or none, fluviometric stations, through the determination of homogeneous regions of minimum flows, using the k-means cluster analysis methodology, enable a better management of water resources in the state of Goiás in Brazil.

MATERIALS AND METHODS

Study area

The study was conducted for the state of Goiás, which is in the Center-Western region of Brazil. It occupies an area of 340,242.82 km² (IBGE, 2020), which represents about 4% of the national territory. Goiás has 246 municipalities and an estimated population of approximately 7.2 million inhabitants (IBGE, 2021). The study area is divided into 11 watersheds according to the State Water Resources Plan -PERH (Goiás, 2016), considered as Water Resources Planning and Management Units (WRPMU/UPGRH). The low density of fluviometric stations located in the southern and northern regions of the state can also be observed in Figure 1.

Methods

With the information provided by the National Water and Sanitation Agency (ANA), through HIDROWEB system, 71 fluviometric stations were used (Figure 1) with at least 5 years of continuously observed data. In this way, the number of stations is below the minimum recommended quantity (182). This recommendation is based on the indications of the WMO (2008) of the need for a station every 1,875 km², taking into consideration inland plains and undulating reliefs. As proposed by Tucci (2001), in the regions with well determined dry and wet seasons, as is the case of Goiás, it is possible to perform the hydrological year approach. Because of this, the seasons were organized into hydrological years, in which the month of March is considered as the beginning of the hydrological year, according to Honorio (2020).

With the historical series organized, the permanence curves for each station were determined.

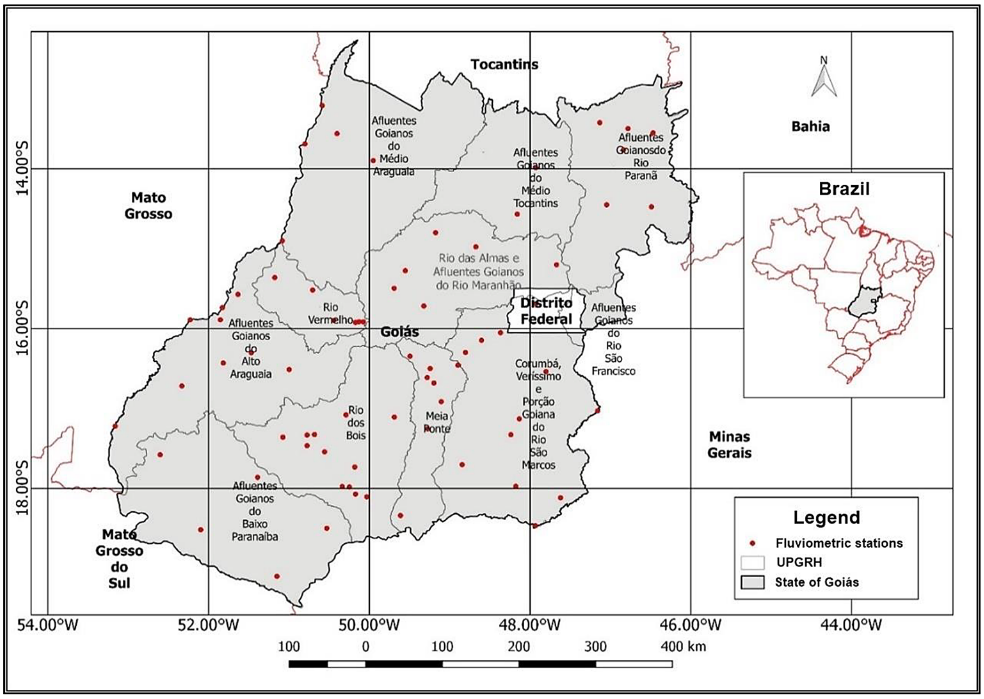

Figure 1. State of Goiás with the division into UPGRH and arrangement of the fluviometric stations

From this, the flow rate that is matched or exceeded during 90 percent of the time (Q90) and the one that is matched or exceeded during 95 percent of the time (Q95) were obtained. These two flows (Q90 and Q95) are used in many states as a grantable reference flow. However, in other states, the reference flow is the minimum flow of 7 days duration and 10 years of recurrence time (Q7, 10). Yet, for the determination of Q7,10, at least 10 years of observed data are necessary. Although 11 of the 71 selected stations did not meet this requirement, they could not be discarded as they belong to basins with little information. Thus, to determine Q7.10 in these 11 stations, the Log-Pearson type III probability distribution was used, which best adhered to the data, according to the Kolmogorov-Smirnov test.

Therefore, with the reference flows (Q90, Q95 and Q7.10) of the fluviometric stations, to better characterize the hydrographic basins of the state, the physical and geological descriptions that present greater impacts on the behavior of the flow of BH were carried out. In this way, the BH drainage area and perimeter, the main variables in determining the water potential, were determined. Regarding the verification of the possibility of recharge of the water table and infiltration (important for minimum flow), the BH were characterized in relation to the average slope, circularity and compactness index, type and use of the soil.

The method of the curve number (CN) was used for the characterization of the type and use of the soil.In this method –presented by the Soil Conservation Service (SCS) in mid-1950s –characteristic values for each type and use of the soil are shown in a standard table of the SCS (Araújo-Neto et al., 2012). The determination of the weighted CN of each watershed was carried out using MapBiomas data (Souza et al. 2020), which presents the products of type and land use. For the grouping to have greater accuracy, the minimum flows were transformed into specific flows, directly relating the area of the hydrographic basin. To explain the shape of the basin, the compactness and circularity indices were used, obtained through the relationship between the area and the perimeter.

Data processing

Initially, a linear correlation analysis was performed at a significance level of 95%, among all hydrological variables. This analysis aimed to identify similar behaviors among the variables, previously indicating the main variables to be used in the application of the cluster analysis. In this way, the lowest intra-group variability and greater variability between the groups was verified and observed..Cluster analysis was performed with all stations and all variables previously characterized. The grouping of stations was performed with the help of the Statistica 7.0 software, which allows the analysis using hierarchical and non-hierarchical methods. The first consists of the mathematical treatment of each sample as a point in the multidimensional space described by the chosen variables. Next, the distance from this point to all the others is calculated, thus defining a descriptive matrix of proximity between all stations.A dendrogram was constructed from the matrix, in which the samples were partitioned into internal homogeneous and external heterogeneous groups. The partition can be performed with different algorithms, such as single-linkage, complete-linkage, and Ward method, among others (Hosking and Wallis, 1997).

The non-hierarchical K-means agglomerative method aims to partition the data into several "K" clusters in order to minimize the internal distance and maximize the external distance between them. To do so, it uses the Euclidean distance as a measure of dissimilarity between the variables (Borges, 2010). This method was used in data grouping because it is considered simple and presents good results.The Euclidean distance was obtained using the Pythagorean theorem for a multidimensional space. It is a measure of dissimilarity that corresponds to the square root of the sum of the squares of the differences between the pairs of observations (a and b) for all n variables (Bussab et al., 1990), according to Eq. 1. The closer to 0, the greater the degree of similarity of the objects of study (Seidel et al., 2008).

$$d_{ab} = \sqrt{\sum_{k=1}^{n}}(X_{ak} - X_{bk})^2 \qquad (1)$$

where: d_{ab} – the distance between elements a and b;
X_{ak} – the value assumed by the variable k of element a;
X_{bk} – the value assumed by the variable k of element b.

First, K initial centroids are chosen, which represent the centers of the K clusters given by C1, ..., CK, where K ≥ 2. Next, each matrix value is labeled in relation to the centroid of the most similar class. Therefore, the centroids have their values updated based on the values that currently belong to the respective clusters. Thus, the process is repeated as long as the stopping criterion is not obtained (Seidel et al., 2008). In this work, the algorithm ended its execution when the centroids remained unchanged between two iterations.

RESULTS AND DISCUSSION

From the characterization of each watershed in relation to several variables (Table 1), it was possible to observe that the CN ranged from 25 to 88, the flow Q7.10 ranged from 0.00004 to 0.01596 $m^3/s \cdot km^2$, the flow Q90 from 0.00021 to 0.02053 $m^3/s \cdot km^2$, and the flow Q95 between 0.00012 and 0.01684 $m^3/s \cdot km^2$. The average slopes of the stations watersheds were obtained based on the Shuttle Radar Topography Mission (SRTM) relief data and their values ranged from 2.654 to 14.055 degrees, the compactness index ranged from 1.510 to 3.548 and the circularity index ranged from 0.078 to 0.432.

Therefore, the correlation matrix between the variables used in the formation of the clusters (presented in Table 2) was applied. A high correlation is observed between the flows, with values above 0.78, which does not occur for the other variables. An inverse relationship between the flows and the average slope since the correlations ranging between -0.3204 and -0.1659 was also identified. As expected, the correlation between the compactness and roundness indices was considerable, above 0.92, since both are obtained through the relationship between perimeter and area; however, this correlation is inversely proportional.. The relationships of the flows with the indices found and the slope were not relevant because they presented small values, approximately 0.25 and 0.35, respectively.

Several centroids were tested in the application of the k-means algorithm. The cluster that presented the best measures of distance between the groups was obtained using 5 centroids in most scenarios since, initially, all variables were included in the analysis, which resulted in a complex cluster and made it impossible to determine regions. Then, it was grouped based only on the reference flows (Q90, Q95 and Q7.10), which was already expected since the variables present a strong correlation (>0.7). This corroborates the studies by Elesbon et al. (2015), in which they performed a multivariate analysis and indicated that the main grouping variables are the minimum flows and the watershed area. These are directly related to the reference flows, thus obtaining a satisfactory determination of the homogeneous regions.

When analyzing Table 3, it is possible to visualize the description of the behavior of the variables in each of the 5 regions. Formed by 26 stations, region 1 is characterized by values of specific minimum flows of reference with an average of

Table 1. Specific permanence flows and physical characteristics of each fluviometric station, Goiás, Brazil

Code	CN	$Q_{7,10}$ ($m^3/s.km^2$)	Q_{90} ($m^3/s.km^2$)	Q_{95} ($m^3/s.km^2$)	Average slope	Compactness index	Circularity index
60680000	50	0.00224	0.00454	0.00362	4.934	2.595	0.146
60665000	79	0.00285	0.00496	0.00398	4.786	1.930	0.264
60654000	25	0.00303	0.00562	0.00465	5.801	1.935	0.263
60650000	79	0.00289	0.00529	0.00452	5.452	2.001	0.246
60640000	88	0.00216	0.00426	0.00356	5.626	1.984	0.250
60642000	25	0.00268	0.00521	0.00432	5.869	2.090	0.226
60653000	47	0.00260	0.00633	0.00477	4.357	1.518	0.428
60635000	64	0.00290	0.00490	0.00404	6.607	1.609	0.381
60477300	81	0.00169	0.00333	0.00275	3.576	1.923	0.266
60500000	81	0.00291	0.00510	0.00420	4.444	2.456	0.163
60540000	68	0.00252	0.00513	0.00419	4.798	2.045	0.236
60432000	81	0.00628	0.00979	0.00838	4.484	1.949	0.259
60433000	75	0.00261	0.00471	0.00388	6.284	2.054	0.234
60545000	25	0.00312	0.00575	0.00478	5.224	2.205	0.203
60590000	75	0.00179	0.00385	0.00295	4.382	2.372	0.175
60430000	79	0.00456	0.00657	0.00574	6.244	1.974	0.253
60160080	47	0.00333	0.00555	0.00512	4.945	3.295	0.091

Table 1. Cont. Specific permanence flows and physical characteristics of each fluviometric station, Goiás, Brazil

60200000	68	0.00269	0.00480	0.00405	5.093	1.857	0.286
60050000	50	0.00244	0.00446	0.00381	3.748	2.586	0.147
60020000	64	0.00205	0.00429	0.00331	3.346	2.196	0.204
60765000	50	0.00115	0.00261	0.00195	3.278	2.281	0.189
60750000	47	0.00102	0.00237	0.00166	3.925	2.211	0.202
60774000	50	0.00664	0.00976	0.00856	2.985	2.291	0.188
60810000	50	0.00246	0.00549	0.00422	2.654	1.892	0.275
60772000	84	0.00180	0.00339	0.00284	3.828	1.910	0.270
60785005	79	0.00827	0.01055	0.00956	2.932	3.057	0.105
60778000	81	0.00699	0.00989	0.00879	2.816	3.249	0.093
60798000	50	0.00510	0.00770	0.00675	2.872	2.005	0.245
60805000	88	0.00226	0.00445	0.00351	3.414	2.069	0.230
60790000	50	0.00558	0.00790	0.00706	2.886	2.045	0.236
60781000	55	0.00574	0.00864	0.00760	2.822	2.115	0.220
60715000	72	0.00102	0.00257	0.00175	4.894	2.315	0.184
25800000	86	0.00007	0.00058	0.00034	3.570	2.188	0.206
25200000	57	0.00306	0.00448	0.00398	4.855	3.176	0.098
25750000	55	0.00004	0.00021	0.00012	4.792	1.997	0.247
25700000	88	0.00213	0.00391	0.00333	4.506	3.390	0.086
25950000	88	0.00206	0.00338	0.00285	4.266	3.548	0.078
20250000	86	0.00193	0.00422	0.00336	5.729	2.076	0.229
20100000	75	0.00199	0.00417	0.00339	7.097	2.013	0.243
20050000	81	0.00305	0.00419	0.00371	5.149	2.098	0.224
20200000	72	0.00181	0.00389	0.00295	5.328	2.008	0.245
20009000	81	0.00391	0.00618	0.00577	5.951	1.616	0.377
20489100	81	0.00109	0.00255	0.00198	5.963	2.096	0.224
24070000	59	0.00733	0.00889	0.00837	4.017	2.060	0.232
24196000	68	0.00146	0.00224	0.00189	5.249	2.248	0.195
24800000	25	0.00185	0.00319	0.00271	5.527	2.122	0.219
24750000	59	0.00203	0.00385	0.00309	6.263	1.990	0.249
24780000	72	0.00065	0.00220	0.00153	4.990	1.981	0.251
24700000	57	0.00419	0.00557	0.00500	5.221	2.910	0.116
24900000	64	0.00353	0.00524	0.00465	5.033	1.780	0.311
24950000	25	0.00177	0.00319	0.00266	5.107	2.512	0.156
24850000	25	0.00385	0.00513	0.00462	5.256	2.975	0.111
60895000	50	0.00867	0.01136	0.01029	2.682	2.195	0.205
60910000	67	0.00544	0.00763	0.00654	3.506	2.149	0.213
60940000	83	0.01443	0.01695	0.01572	2.804	2.909	0.116
60950000	72	0.00912	0.01141	0.01051	3.175	3.450	0.083
60870000	55	0.00613	0.00880	0.00779	4.047	1.819	0.298
21300000	88	0.00557	0.00687	0.00648	4.889	2.273	0.191
21580000	81	0.01596	0.01733	0.01684	5.836	2.054	0.234
21560000	25	0.00704	0.00849	0.00793	4.580	2.109	0.222
21600000	47	0.00270	0.00326	0.00299	5.175	2.544	0.152
21500000	25	0.00217	0.00275	0.00251	4.914	2.404	0.171
21220000	86	0.00040	0.00091	0.00070	5.581	2.160	0.211
25070000	75	0.00041	0.02053	0.00143	8.085	1.510	0.432
25090000	75	0.00081	0.00233	0.00148	9.567	1.540	0.415
25140000	83	0.00173	0.00269	0.00222	2.969	1.956	0.258
25120000	75	0.00061	0.00198	0.00128	5.894	1.786	0.309
25130000	25	0.00034	0.00140	0.00098	5.531	1.909	0.270
25100000	81	0.00082	0.00264	0.00163	8.889	1.652	0.361
20699000	81	0.00292	0.00442	0.00393	14.055	2.159	0.211
20950000	67	0.00159	0.00195	0.00180	6.533	2.473	0.161

Table 2. Correlation matrix between the variables used in the formation of clusters

Variables	CN	$Q_{7,10}$	Q_{90}	Q_{95}	Average slope	Compactness index	Circularity index
CN	1.0000	0.0358	0.0611	0.0246	0.1114	-0.0167	0.0810
$Q_{7.10}$	0.0358	1.0000	0.7890	0.9899	-0.3048	0.2615	-0.2558
Q_{90}	0.0611	0.7890	1.0000	0.8073	-0.1659	0.0815	0.0160
Q_{95}	0.0246	0.9899	0.8073	1.0000	-0.3204	0.2375	-0.2223
Average slope	0.1114	-0.3048	-0.1659	-0.3204	1.0000	-0.2964	0.3565
Compactness index	-0.0167	0.2615	0.0815	0.2375	-0.2964	1.0000	-0.9238
Circularity index	0.0810	-0.2558	0.0160	-0.2223	0.3565	-0.9238	1.0000

0.00283 for Q7.10, 0.00450 for Q90 and 0.00385 m³/s for Q95. The average slope of this region varies from 4,266 to 5,328 degrees. Region 2, consisting of 19 stations, has average values for the minimum specific flows Q7.10, Q90 and Q95, 0.00278, 0.00551 and 0.00389 m³/s, respectively. Ranging from 5.452 to 8.085 degrees, it is possible to observe the average slope.With the average slope values ranging from 8.889 to 14.055 degrees, 3 stations made up region 3. With this, it was possible to observe the minimum flows with an average of 0.00151 m³/s for Q7.10, 0.00313 m³/s for Q90 and 0.00163 m³/s for Q95.

With an average of 0.00298 for Q7.10, 0.00495 for Q90 and 0.00415 m³/s for Q95, region 4 consists of 11 stations. This region is defined by values of minimum average slope of 3.346 and maximum of 4.357 degrees.Region 5, with slopes ranging from 2.654 to 3.278 degrees, is made up of 12 stations. The average flow Q7.10 is 0.00632 m³/s, with an average of 0.00874 m³/s the flow Q90 is observed. Q95 flow has an average of 0.00776 m³/s.

It can be seen in Table 3 that the regions with the steepest average slope tended to have lower values of reference flows (m³/s.km²), since the minimum flows are maintained by the water table, and the higher the slope, the greater the vertical distance to the surface, in accordance with the results obtained by Nogueira (2017). The same applies for the regions with lower values, that is, the lower the vertical distance, the higher the observed minimum flow rate.The values of the compactness and circularity indices presented similar mean values, as shown in Table 3. These did not interfere considerably in the determination of the 5 regions, as well as the CN values.

Table 4 shows the values obtained for the Euclidean distance (application of equations 1) between the clusters obtained via Statistica 7.0. The distances ranged from 0 (when it comes to the distance from the cluster to itself) to 3.249 (as the distance from cluster 3 to cluster 5), making it possible to analyze that C3 is the one with the highest values of distance to the others.

Figure 2 presents the distribution of the 5 homogeneous regions for the state of Goiás, thus facilitating the visualization of regions with

Table 3. Minimum, average and maximum values of each variable in each homogeneous region

	C1	C2	C3	C4	C5
CN					
Minimum	25	25	75	47	50
Average	61	66	79	66	63
Maximum	88	88	81	88	83
$Q_{7,10}$ (m³/s.km²)					
Minimum	0.00004	0.00034	0.00080	0.00007	0.00115
Average	0.00283	0.00278	0.00151	0.00298	0.00632
Maximum	0.00704	0.01590	0.00291	0.00733	0.01442
Q_{90}(m³/s.km²)					
Minimum	0.00209	0.00091	0.00232	0.00058	0.00261
Average	0.00450	0.00551	0.00313	0.00495	0.00874
Maximum	0.00978	0.02054	0.00442	0.00889	0.01694
Q_{95}(m³/s.km²)					
Minimum	0.00012	0.00070	0.00148	0.00033	0.00195
Average	0.00385	0.00389	0.00163	0.00415	0.00776
Maximum	0.00838	0.01684	0.00392	0.00836	0.01571
Average Slope (°)					
Minimum	4.266	5.452	8.889	3.346	2.654
Average	4.859	6.098	10.837	3.756	2.906
Maximum	5.328	8.085	14.055	4.357	3.278
Compactness Index					
Minimum	1.780	1.510	1.540	1.517	1.892
Average	2.421	1.971	1.783	2.057	2.453
Maximum	3.548	2.473	2.159	2.585	3.449
Circularity Index					
Minimum	0.078	0.161	0.211	0.147	0.082
Average	0.186	0.263	0.329	0.245	0.184
Maximum	0.311	0.432	0.415	0.427	0.275

Table 4. Euclidean distance (below the diagonal) and Euclidean distance squared (above the diagonal) between the formed clusters

	C1	C2	C3	C4	C5
C1	0.000	0.276	5.954	0.239	0.659
C2	0.525	0.000	3.749	0.915	1.738
C3	2.440	1.936	0.000	8.366	10.561
C4	0.488	0.956	2.893	0.000	0.147
C5	0.812	1.318	3.249	0.384	0.000

homogeneous hydrological behavior. Hence, it is possible to see that the central region of the state is mostly composed of region 2, the east and west ends represented by region 1, to the south, a mix composed of region 4 and 5. Region 3 is difficult to observe, since it only comprises 3 stations constituted it, highlighting how important it is to expand monitoring in these regions to increase the reliability of the clusters. The northernmost and southernmost portions were not considered, as they do not have stations analyzed in this study. Thus, it is possible to observe that the methodology presented here indicated that the 11 UPGRH (Water Resources Planning and Management Units) in the state of Goiás could be shortened to 5 regions. This would simplify the control and inspection of allocation processes and would also improve the issue of accuracy regarding the reference flows presented in the State Plan for Water Resources-PERH (Goiás, 2016). It is noteworthy that the regions without information were not included in any other region due to the uncertainties involved in these areas.

Therefore, it is noteworthy that fluviometric monitoring in Goiás is more uniform in the central region than in the extremities of the state, where there are few monitoring stations. This fact indicates that there is a need to improve the distribution of fluviometric monitoring station. In order to occur more accurate technical studies, it is necessary to expand monitoring in the state of Goiás and in Brazil as a whole. The availability of correct data to characterize the different watersheds throughout the state of Goiás is a crucial instrument to correctly manage the state´s water resources and thus conserve this precious good to present and future generations and avoid water scarcity and crisis situations.Thus, with the results presented here, there is an indication that the use of physical and hydrological characteristics of watershed improves performance and optimizes the regionalization of minimum flows. When using fluviometric stations from different basins with different characteristics, they often present estimates of water availability in small basins, but when observing the monitoring the water course dries up in the dry season. It is noteworthy that fluviometric monitoring in Goiás is more uniform in the central region, while at the extremes of the state there are few monitoring stations. There is a need to improve the distribution of fluviometric monitoring stations for setting up more accurate technical studies.

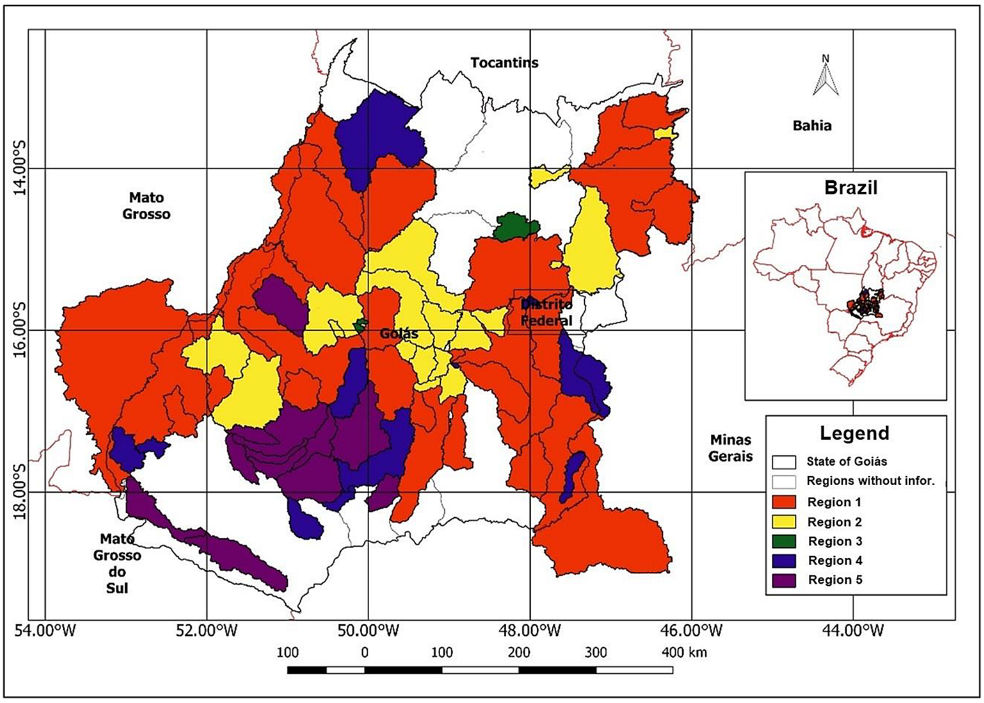

Figure 2. Regions with homogeneous behavior in the state of Goiás

CONCLUSIONS

The methodology presented in this study made it possible to indicate regions and planning units from the fluviometric stations in the State of Goiás (Brazil). The grouping that used the cluster method together with the reference flows, based on stations and their hydrographic basins, showed a great accuracy, as it considered the observed flow and the local characteristics, rather than the geographic region in which the station was located. Therefore, five specific homogeneous regions were determined, which enabled obtaining minimum flows for each region,, indicating the reference flows and their characteristics. It is also noteworthy that the use of the watershed average slope was the physical characteristic that most influenced the definition of homogeneous regions. Therefore, the methodology presented here provided a robust grouping in relation to the minimum reference flows and it can be implemented in other states/regions, thus improving the quality of the estimation of local reference flows.

Acknowledgements

We thank the National Health Foundation (FUNASA) of Brazil for their financial support, through the project entitled: Sanitation and Environmental Health in Rural and Traditional Communities in Goiás (SanRural) - TED 05/2017.This research also had the support of the project UIDB/00195/2020 (FibEnTech) and UIDB/04035/2020 (GeoBioTec), both funded by the Fundação para a Ciencia e Tecnologia (FCT) in Portugal.

REFERENCES

1. Araújo-Neto, J.R.; Palácio, H.A.Q.; Andrade, E.M.; Santos, J.C.N.; Pinheiro, E.A.R., 2012. Optimization of the cuver number (Cn-Scs) for different manajegement in a semarid region, Ceará, Brasil. Revista IRRIGA – Brazilian Journal of Irrigation and Drainage, Edição Especial,v. 1, (1), 264-279. https://doi.org/10.15809/irriga.2012v1n01p264
2. Araújo, A. P.; Rocha, P. C. Regime de fluxo e alterações hidrológicas no rio Tibagi-bacia do rio Paranapanema / Alto Paraná. Revista de Geografia, v. 7, n. 3, 2010.
3. Aytaç, E., 2020. Unsupervised learning approach in defining the similarity of catchments: Hydrological response unit based k-means clustering, a demonstration on Western Black Sea Region of Turkey. International Soil and Water Conservation Research v. 8, (3), 321-331. https://doi.org/10.1016/j.iswcr.2020.05.002
4. Basso, R.; Honório, M.; Costa, I.; Bezerra, N.; Baumann, L.; Silva, F.; Albuquerque, A.; Scalize, P. Comparison between Regionalized Minimum Reference Flow and On-Site Measurements in Hydrographic Basins of Rural Communities in the State of Goiás, Brazil. Water 2022, 14, 1016. https://doi.org/10.3390/w14071016
5. Beskow, S.; Norton, L.; Mello, C. Hydrological prediction in a tropical watershed dominated by Oxisols using a distributed hydrological model. Water Resour. Manag. 2013, 27, 341–363.
6. Borges, V.R.P. Comparação entre as técnicas de agrupamento K-Means e Fuzzy C-Means para Segmentação de Imagens Coloridas. VIII Encontro Anual de Computação Out (2010). (Accessed Jul., 2020) at: https://www.enacomp.com.br/2010/anais/artigos/completos/enacomp2010_42.pdf
7. 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 e dá outras providências. Diário Oficial da União: seção 1. Brasília – DF, p. 470-474, 9 jan. 1997. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l9433.htm. Acesso em: 15 out. 2020.
8. Burn, D.H., 1989. Cluster analysis as applied to regional flood frequency. Journal of Water Resources Planning and Management, v. 115, (5), 567-582. https://doi.org/10.1061/(ASCE)0733-9496(1989)115:5(567)
9. Bussab, W.O.; Miazak, E.S.; Andrade, D.F., 1990. Introdução à análise de agrupamentos. 9º Simpósio Brasileiro de Probabilidade e Estatística. São Paulo: IME – USP. (Accessed Feb 1, 2022) at https://cicerocq.files.wordpress.com/2019/05/analise-de-agrupamento_bussab.pdf
10. Cupak, A.; Walega, A.; Michalec, B., 2017. Cluster analysis in determination of hydrologically homogeneous regions with low flow. Acta Scientiarum Polonorum Formatio Circumiectus, v. 16, (1), 53–63. http://dx.doi.org/10.15576/ASP.FC/2017.16.1.53
11. Demirel, M.C.; Mariano, A.J.; Kahya, E., 2007. Performing k-means analysis to drought principal components of Turkish rivers. Hydrology Days, v. 1, 145-151. http://dx.doi.org/10.25675/10217/200671
12. Dikbas, F., Firat, M., Koc, A.C.; Gungor, M., 2013. Defining homogeneous regions for streamflow processes in Turkey using a K-means clustering method. Arabian Journal for Science and Engineering, v. 38, 1313–1319. https://doi.org/10.1007/s13369-013-0542-0
13. Elesbon, A.A.A.; Silva, D.D.; Sediyama, G.C.; Guedes, H.A.S.; Ribeiro, C.A.A.S.; Ribeiro, C.B.

M. Multivariate statistical analysis to support the minimum stream flow regionalization. EngenhariaAgrícola, v. 35, n. 5, 2015, 838-851 p

14. Eslamian, S.; Biabanaki, M. Low flow regionalization modeling. International Journal of Ecological Economics & Statistics, v. 12, n.F08,p. 82-97, 2008.
15. Farsadnia, F.; Kamrood, M.S.; Nia, A.M.; Modarres, R.; Bray, M.T.; Han, D.; Sadatinejad, J., 2014. Identification of homogeneous regions for regionalization of watersheds by two-level self-organizing feature maps. Journal of Hydrology, v. 509, 387–397. https://doi.org/10.1016/j.jhydrol.2013.11.050
16. Gaál, L.; Szolgay, J.; Lapin, M.; Fasko, P., 2009. Hybrid approach to delineation of homogeneous regions for regional precipitation frequency analysis. *Journal of Hydrology and Hydromechanics,* v. *57,* (4), 226-249. https://doi.org/10.2478/v10098-009-0021-1
17. Goiás (2019). Estabelece o regulamento do sistema de outorga das águas de domínio do estado de Goiás e dá outras providências. Conselho Estadual de Recursos Hídricos. Resolução nº 22 de 2019, Goiás, Brazil.
18. Goiás (2019). Secretaria de Estado de Meio Ambiente e Desenvolvimento Sustentável. Resolução nº 09 de 2005. Estabelece o Regulamento do Sistema de outorga das águas de domínio do Estado de Goiás e dá outras providências. Publicado no Diário Oficial do Estado de Goiás n.º 19.653, de 02/06/2005, Goiás, Brazil.
19. Goiás (2016). Plano Estadual de Recursos Hídricos do Estado de Goiás. Secretaria de Estado de Meio Ambiente, Recursos Hídricos, Infraestrutura, Cidades e Assuntos Metropolitanos, 2016, SECIMA, Goiás, Brazil.
20. Gomes, E.P.; Blanco, C.J.C.; Pessoa, F.C.L., 2018. Identification of homogeneous precipitation regions via Fuzzy c-means in the hydrographic region of Tocantins–Araguaia of Brazilian Amazonia. Applied Water Science, v. 9, (6), 1-12. https://doi.org/10.1007/s13201-018-0884-6
21. Granemann, A. R. B., Mine, M. R. M., Kaviski, E. Frequency analysis of minimum flows. Brazilian, Journal of Water Resources, v. 23, e17, p. 1–14, 2018.
22. Hall, M.J.; Minns, A.W., 1999.The classification of hydrologically homogeneous regions. Hydrological Sciences Journal, v. 44, (5), 693-704. https://doi.org/10.1080/02626669909492268
23. Honório, M. S. Avaliação da disponibilidade hídrica superficial no estado de Goiás. Dissertação (Mestrado em Ciências Ambientais) - Universidade Federal de Goiás, Pró Reitoria de Pós-graduação (PRPG), Goiânia, 2020.
24. Hosking, J.; Wallis, J., 1997. Identification of homogeneus regions. Regional Frequency Analysis: An Approach Based on L-Moments. Cambridge: Cambridge University Press, 54-72. https://doi.org/10.1017/CBO9780511529443
25. IBGE, 2020. Área territorial Brasileira. Instituto Brasileiro de Geografia e Estatística, (Accessed Feb 1, 2022) at: https://cidades.ibge.gov.br/brasil/go/panorama
26. IBGE, 2021. Diretoria de Pesquisas, Coordenação de População e Indicadores Sociais, Estimativas da população residente com data de referência 1o de julho de 2021. Instituto Brasileiro de Geografia e Estatística (Accessed Feb 1, 2022) at: https://cidades.ibge.gov.br/brasil/go/panorama
27. Kahya, E.; Demirel, M.C.; Piechota, T.C., 2007. Spatial grouping of annual streamflow patterns in Turkey. Hydrology Days, v.1, 169-176. http://dx.doi.org/10.13140/RG.2.1.2999.7849
28. Kazemzadeh, M.; Malekian, A., 2018. Homogeneity analysis of streamflow records in arid and semiarid regions of northwestern Iran. Journal of Arid Land, v. 10, (4), 493–506. https://doi.org/10.1007/s40333-018-0064-4
29. Kim, D.; Jung, I.; Chun, J.A., 2016. A comparison between parameter regionalization and model calibration with flow duration curves for prediction in ungauged catchments. Hydrology and Earth System Sciences Discussions, v. 5, (8), 1-29. http://dx.doi.org/10.5194/hess-2016-487
30. Lücke, J.; Forster, D., 2019. k-means as a variational EM approximation of Gaussian mixture models. Pattern Recognition Letters, v. 125, 349-356. https://doi.org/10.1016/j.patrec.2019.04.001.
31. Maciel, A. L.; Vieira, E. M.; Monte Mor, R. C.; Vasques, A. C. Regionalização e espacialização de vazões de permanência: Estudo aplicado na bacia do rio Piracicaba - MG. Revista Brasileira de Climatologia, v. 24, p. 114–133, 2019.
32. Nadoushani, S.S.M.; Dehghanian, N.; Saghafian, B., 2018. A fuzzy hybrid clustering method for identifying hydrologic homogeneous regions. Journal of Hydroinformatics, v. 20, (6), 1367–1386. https://doi.org/10.2166/hydro.2018.004
33. Nathan, R.J.; McMahon, T.A., 1990. Identification of homogeneous regions for the purposes of regionalisation. Journal of Hydrology, v. 121, 217-238. Disponível em: https://doi.org/10.1016/0022-1694(90)90233-N
34. Nogueira, S. H. M. Avaliação do risco à perda da qualidade ambiental do aquífero freático na região metropolitana de Goiânia. Dissertação de Mestrado. Programa de Pós-graduação em Ciências do Ambiente, UFG, Goiânia. 2017.
35. Pruski, F. F.; Nunes, A. A.; Rego, F. S.; Souza, M. F. Extrapolação de equações de regionalização de vazões mínimas: Alternativas para atenuar os riscos. Water Resources and Irrigation Management, v. 1, n. 1, p. 51-59, 2012. doi.org/10.3390/rs12172735

36. Razavi, T.; Coulibaly, P., 2013. Streamflow Prediction in Ungauged Basins: Review of Regionalization Methods. Journal of Hydrologic Engineering, v. 18, 958-975. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000690
37. Samuel, J.; Coulibaly, P.; Metcalfe, R. A. Estimation of continuous streamflow in Ontario ungauged basins: Comparison of regionalization methods. Journal of Hydrologic Engineering, v. 16, p.447, 2011.
38. Seidel, E.J.; Moreira Júnior, F.J.; Ansuj, A.P.; Noal, M.R.C., 2008. Comparação entre o método Ward e o método K-médias no agrupamento de produtores de leite. Ciência e Natura, v. 30, (1), 7- 15. https://doi.org/10.5902/2179460X9737
39. Silva, A. C., 2018. Identificação de regiões hidrologicamente homogêneas por agrupamento fuzzy c-means no estado do Paraná. Dissertação de Mestrado, Programa de Pós-Graduação em Engenharia Agrícola, UNIOESTE, Cascavel – Paraná. (Accessed Feb 1, 2022) at: http://131.255.84.103/bitstream/tede/3760/5/Ana%20Claudia_Silva2018.pdf.
40. Smakhtin, V. U. Low flow hydrology: a review. Journal of hydrology, v. 240, n. 3-4, p. 147-186, 2001.
41. Souza, Jr. C.M.; Shimbo, J.Z; Rosa, M.R; Parente, L.L.; Alencar, A.A; Rudorff, B.F.T.; Hasenack, H.; Matsumoto, M.; Ferreira, L.G; Souza-Filho, P.W.M.; Oliveira, S.W.; Rocha, W.F.; Monteiro, D.; Rosa, E.R.; Vélez-Martin, C.; Weber, E.J.; Lenti, F.E.B.; Paternost, F.F.; Pareyn, F.G.C.; Siqueira, J.V.; Viera, J.L.; Ferreira Neto, L.C.; Saraiva, M.M.; Sales, M.H.; Salgado, M.P.G.; Vasconcelos, R.; Galano, S.; Mesquita, V.V.; Azevedo, T., 2020. Reconstructing three decades of land use and land cover changes in brazilian biomes with landsat archive and earth engine. Remote Sensing, v.12, Issue 17. https://doi.org/10.3390/rs12172735
42. Tucci, C.; Silveira, A.; Sanchez, J.; Albuquerque, F., 1995. Flow regionalization in the upper Paraguay basin, Brazil. Hydrological Sciences Journal, v. 40, (4), 485-497. https://doi.org/10.1080/02626669509491434
43. Tucci, C.E.M., 2001. Regionalização de vazões. In: Tucci, C.E.M. Hidrologia: ciência e aplicação. 2. ed. Porto Alegre: Editora da UFRGS, p. 573-611.
44. WMO (2008). Guide to Hydrological Practice Hydrology – From Measurement to Hydrological Information. Sixth edition. World Meteorological Organization, Geneva, Switzerland, v. 1, (168), p.296. (Accessed Feb 1, 2022) at: https://library.wmo.int/doc_num.php?explnum_id=10473
45. Wu, H.Q., Huang, Q., Xu, W., Xi, S.F., 2015. Application of K-Means Cluster and rough set in classified real-time flood forecasting. Advanced Materials Research, v. 1092–1093, 734–741. https://doi.org/10.4028/www.scientific.net/amr.1092-1093.734.
46. Zhang, Z.; Stadnyk, T.A., 2020. Investigation of attributes for identifying homogeneous flood regions for regional flood frequency analysis in Canada. *Water, v. 12, (9),* 1-24. https://doi.org/10.3390/w12092570