

Doi: <http://dx.doi.org/10.1590/1809-4430-Eng.Agric.v40n3p334-343/2020>

## PHYSICAL ANALYSIS OF REGIONALIZED FLOW AS AN AID IN THE IDENTIFICATION OF HYDROLOGICALLY HOMOGENEOUS REGIONS

Arthur T. Calegario<sup>1\*</sup>, Fernando F. Pruski<sup>2</sup>, Rayssa B. Ribeiro<sup>1</sup>, Maria C. A. Ramos<sup>1</sup>,  
Fernando S. Rego<sup>3</sup>

<sup>1\*</sup>Corresponding author. Federal University of Viçosa (UFV)/ Viçosa - MG, Brazil.  
E-mail: [tcalegario@hotmail.com](mailto:tcalegario@hotmail.com) | ORCID ID: <https://orcid.org/0000-0002-1921-5234>

### KEYWORDS

regionalization, long-term average streamflow, planning and management of water resources, cluster analysis, geographical convenience.

### ABSTRACT

Regionalization is an important technique for estimating the flow of hydrographic sections with a lack of data. First, it is necessary to identify hydrologically homogeneous regions (HHRs), which are commonly validated via statistical analyses. Because this step is understood to be subjective, studies that contribute to a greater reliability in identifying regions are needed. In this context, the objective was to evaluate the inclusion of a physical analysis of the average regionalized flow rates as an aid to identify HHRs. The groupings were defined on the basis of geographical convenience methods and cluster analysis. For the assessment of regionalized flows, six statistical indices were used with a physical analysis that was performed via a comparison of the runoff coefficient to the spatial distribution of precipitation values. It was concluded that the physical analysis reduced the subjectivity in the identification of HHRs.

### INTRODUCTION

The study of flows is the basis for important decisions in the planning and management of water resources (Coxon et al., 2015; Westerberg et al., 2014) and solving environmental and engineering problems, such as sizing structures for water use and control, economic evaluation of flood protection projects, planning and management of land use, water quality control, among others. (Agarwal et al., 2016). The flow data are restricted to places where stream gauge stations are available (Pruski et al., 2016), and not always providing information in regions of interest (Nruthya & Srinivas, 2015). For these reasons, it is necessary the use of tools for prediction of flow (Kim et al., 2016).

The regionalization of flows comprises techniques used to overcome the lack of hydrological data in places where they are scarce or nonexistent (Kult et al., 2014; Beskow et al., 2014, Liew & Mittelstet, 2018). Although there is no standard methodology for regionalization in the literature (Razavi & Coulibaly, 2013), regression models are the most widely used, and consist of estimating flows

based on equations that relate hydrological information of interest to catchment characteristics (Arsenault & Brissette, 2014).

The regionalization of stream gauge indices is based on the premise that basins with a similar climate, geology, topography, vegetation and soils have similar hydrological responses; however, river basins with large drainage areas may have distinct hydrologic behaviors along their hydrography (Smakhtin, 2001). Thus, to ensure greater security and predictive reliability (Nathan & McMahon, 1990; Smakhtin, 2001), a step in the study of flow regionalization consists of dividing the studied area into regions with a similar behavior (Hosking & Wallis, 1997), which are known as hydrologically homogeneous regions (HHRs).

In the literature, many methods to identify HHRs are mentioned, such as cluster analysis, residual standard, seasonality indices, classification and regression trees, canonical correlation analysis, entropy and geographical convenience. However, this is a subjective and difficult step, as there is no agreement on an objective technique for the determination of these regions (Hosking & Wallis,

<sup>1</sup> Universidade Federal de Viçosa (UFV)/ Viçosa - MG, Brasil.

<sup>2</sup> Universidade Federal de Viçosa(UFV)/ Bolsista 1A do CNPq/ Viçosa - MG, Brasil.

<sup>3</sup> Serviço Geológico do Brasil (CPRM)/ Belo Horizonte - MG, Brasil.

Area Editor: Alexandre Barcellos Dalri

Received in: 8-10-2018

Accepted in: 3-24-2020



1997), and its validation is usually based on statistical analyses only (Nathan & McMahon, 1990).

In addition to defining HHRs, another difficulty with respect to regionalization is that the use of regression models is not recommended beyond the limits of the sample data (Naghetini & Pinto, 2007), which in practice means that flow regionalization conducted using stream gauge stations associated with large drainage areas is restricted to an unimportant part of the hydrography, making the planning and management of water resources more difficult (Silva Júnior et al., 2003). In this sense, the procedural association that helps to understand the physical behavior of the process is important for extracting more information from the available data (Pruski et al., 2013; Pruski et al., 2015).

This study was developed under the assumption that the physical behavior of regionalized flows, together with statistical analysis, may help to identify HHRs, lessening the uncertainties associated with flow regionalization. The aim, therefore, was to evaluate the incorporation of a physical analysis into average regionalized flows as an aid to identify HHRs.

## MATERIAL AND METHODS

### Study area and data used

The study region corresponds to part of the Doce River basin in Minas Gerais (Figure 1), which covers an area of 71724 km<sup>2</sup> and represents 86% of the total basin area. The pluviometric regime of the Doce River basin is characterized by two quite distinctive periods: a rainy season from October to March, with the highest rates during December, and a dry season from April to September (Louzada et al., 2018).

In this study, we used data from 1975 to 2005 from 38 stream gauge stations and 80 pluviometric stations that belong to the hydro-meteorological network of the Hydrological Information System (Hidroweb) of the Brazilian National Water Agency (ANA) in addition to data from 1961 to 1990 from 14 climatological stations that are part of the station network of the Brazilian National Institute of Meteorology (INMET). The study stations are presented in Figure 1.

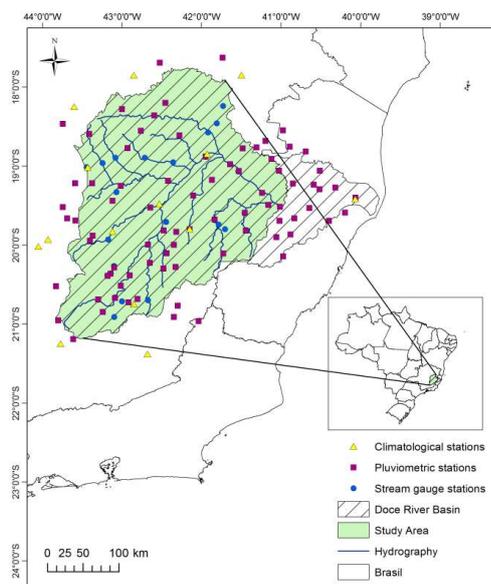


FIGURE 1. Location of the study area and pluviometric, stream gauges and climatological study stations.

### HHR identification

To identify HHRs, we used two methods: geographical convenience and cluster analysis. For this paper, identification of an HHR using geographical convenience was performed considering the basin precipitation map together with statistical analyses of regression models and the behavior of regionalized flows to attempt to reduce delimitation mistakes.

For application of cluster analysis, it is necessary to choose the variables, agglomeration method and dissimilarity measure, all of which have a strong influence on the results. Cluster analysis was conducted using different combinations of latitude, longitude, long-term average rainfall and real evapotranspiration variables to characterize their influence on the identification of HHRs and thus on the regionalized flows. Latitude and longitude were used to obtain geographically continuous regions (Rao & Srinivas, 2006), and the long-term average precipitation and real evapotranspiration were used to represent the input and output, respectively, of water in the hydrographic basin as they are usually the main components of the water balance.

Long-term average precipitation (P) was obtained from data from the analyzed historical series, as this information's spatialization to the basin was applied by interpolation using the IDW (Inverse Distance Weighting) method, as shown in Figure 2a.

Evapotranspiration calculus was conducted using the Penman-Monteith FAO 56 as described by Allen et al. (1998) to obtain the monthly reference evapotranspiration (ET<sub>o</sub>) for each climatological station. Later, ET<sub>o</sub> was interpolated using the IDW method. For the real evapotranspiration (ET<sub>r</sub>) estimation at every pluviometric station, we used the climatological water balance (CWB) method of Thornthwaite & Mather (1955), adopting the available water capacity as equal to 100 mm (Passos et al., 2017). The yearly average ET<sub>r</sub> at the pluviometric station locations was obtained by summing the monthly amounts. By performing interpolation using the IDW method, we obtained a specialized map of the yearly average ET<sub>r</sub> (Figure 2b).

Cluster analysis is dependent on the units and scales by which the descriptive variables used are measured (Nathan & McMahon, 1990); thus, the variables was rescaled base on [eq. (1)].

$$y_{ik} = \frac{x_{ik} - \bar{x}_k}{\sigma_k} \quad \text{for } 1 \leq i \leq n, 1 \leq k \leq N \quad (1)$$

Where:

$n$  = number of individuals;

$N$  = total number of variables used;

$y_{ik}$  = variables 'x<sub>ik</sub>' standardized;

$\sigma_k$  = standard deviation of the variable 'k', and

$\bar{x}_k$  = variable mean 'k'

In the cluster grouping analysis, the non-hierarchical K-means method was used and required a priori the number of groups to be formed. As this value was not known in advance, we used four indices for group validation to estimate the number of groups: Calinski Harabasz (Calinski

& Harabasz, 1974), Dunn (Dunn, 1974), Silhouette width (Rousseeuw, 1987) and Xie-Beni (Xie & Beni, 1991).

The dissimilarity measure adopted in this study was generalized Euclidean distance.

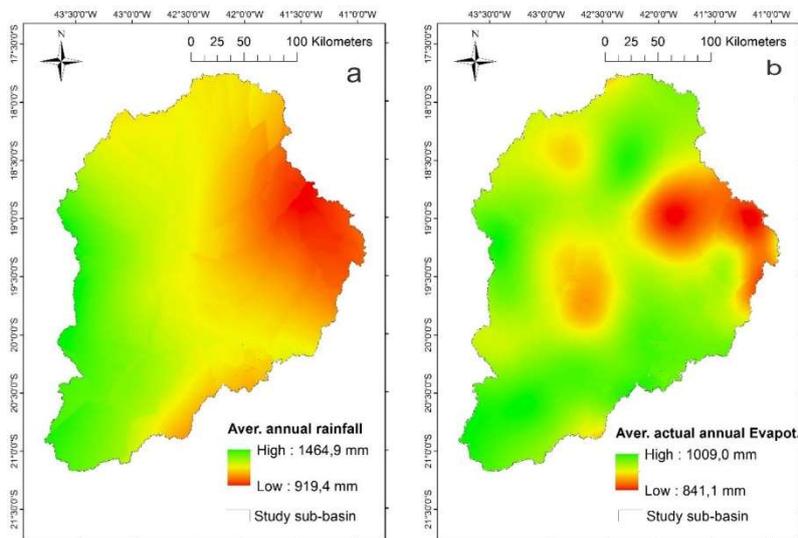


FIGURE 2. Average long-term precipitation (a) and average annual evapotranspiration (b) of the Doce River Basin.

### Flow regionalization

Parametric regression is one of the most widely used methods in regionalization studies (Pruski et al., 2013). To relate the flow to the basin characteristics, it is usually used as a potential function (Samuel et al., 2011).

The dependent variable used in flow regionalization was the long-term average flow ( $Q_{mt}$ ), which represents the potential water availability of the basin. The independent variable used (equation 2) was the flow equivalent to the precipitated volume, considering the subtraction of the precipitation abstraction factor for the flow formation (Novaes, 2005; Pruski et al., 2013; Gonçalves et al., 2018). This factor considers a part of the rain that is not converted to the stream mainly due to evapotranspiration, so [eq. (2)] it is an indirect estimate of the mean runoff in the watercourses. The value of 750 mm was used because demonstrates better results in previously studies in Doce river basin (GPRH & IGAM, 2012).

$$Peq_{750} = \frac{(P-750) A}{31536} \quad (2)$$

Where:

$Peq_{750}$  = equivalent flow of a precipitated volume of less 750 mm of the abstraction in  $m^3 s^{-1}$ ;

$P$  = average annual precipitation in the considered drainage area in mm;

$A$  = drainage area in  $km^2$ ,

31536 = a constant that converts  $P$  to meters,  $A$  to squared meters, and converts the time-step from year to seconds.

The use of  $Peq_{750}$  allows one to relate the precipitation and drainage areas as a single variable that has the advantage of a two-dimensional representation of the relationship between the dependent and independent

variables. In addition, another benefit is obtaining an additional degree of freedom for statistical analysis, reducing the variance of the estimates for an equation with a single explanatory variable compared to an equation with two variables (Pruski et al., 2013).

To obtain the drainage area of the total of each river section, we used the Hydrographically Conditioned Digital Elevation Model (HCDEM) generated from curve graphs by the Brazilian Institute of Geography and Statistics (IBGE) and the mapped hydrography at a scale of 1:100.000 and 1:25.000, when available.

### Statistical behavior analysis of the regionalized flows

An analysis of the regression model fitting to the sample data was conducted for each HHR based on the coefficient of determination ( $R^2$ ). A comparison of the method's efficiency for identifying HHRs from the flow estimations was based on five statistical indices: (a) amplitude of relative error ( $AE_i$ ), which is the largest and smallest relative error value (Pruski et al., 2013); (b) the average relative error ( $ER_m$ ); (c) the Nash-Sutcliffe efficiency with logarithm data ( $E_1$ ) (Krause et al., 2005; Samuel et al., 2011); (d) the relativized Nash-Sutcliffe efficiency ( $E_2$ ) (Krause et al., 2005); and (e) the modified Nash-Sutcliffe efficiency ( $E_3$ ) (Krause et al., 2005).

### Physical behavior analysis of the regionalized flow

Regionalization models obtained for each HHR allow one to estimate the  $Q_{mt}$  along the hydrography. However, performing a physical examination of the estimated absolute flows is a difficult task because of their interval of variation. To address this problem, physical behavior analysis of  $Q_{mt}$  was performed considering the runoff coefficient ( $RC$ ) obtained using the following equation:

$$RC = \frac{Q_{mt}}{Peq} \quad (2)$$

$$Peq = \frac{P A}{31536} \tag{3}$$

Where:

$Peq$  = equivalent flow of a precipitated volume in  $m^3 s^{-1}$ .

The physical behavior evaluation was conducted with an analysis of the  $RC$  spatial distribution by checking the trend of the behavior of this variable in relation to the precipitation map. In addition, the  $RC$  mean values were compared to the precipitation mean and the estimated average  $RC$  with the range observed in the HHRs.

At the same time, a safety analysis associated with the obtained estimates was proposed to mitigate the uncertainty associated with the flow regionalization, especially in extrapolation areas at the limit of the validity of the regression equation.

Safety analysis was performed by means of a security histogram built in two stages. The first refers to the construction of an individual histogram for each of the identified HHRs. Each individual histogram interval was defined on the basis of the minimum and maximum values of the observed  $RC$  in the HHR stream gauge stations in the analysis and considering that the  $RC$  maximum physical limit was 1. It is noteworthy that the  $RC$  may be greater than one if subsurface flows exist between two or more adjacent catchments (Şen & Altunkaynak, 2006), like in Karstic regions, which is not the case for Doce river basin. The second step consisted of the union of the hydrography sections of the different HHRs contained in the same individual histogram intervals. In this way, one can build a final histogram for each grouping.

An estimated  $RC$  value lower than the lowest  $RC$  value observed in the stream gauge stations is physically possible and provides greater security in comparison to the planning and management of water resources, since compared to the values observed in the stream gauge stations, they underestimate the  $RC$ .

The estimated  $RC$  values contained in the interval consisted of the maximum and minimum values for the  $RC$  observed at the stream gauge stations that are physically acceptable and likely.

The estimated  $RC$  values between the maximum value observed in the stream gauge stations and maximum value of the physically possible  $RC$  ( $RC = 1$ ) characterize an interval whose values are considered to be unreliable for the planning and management of water resources, as they may lead to an overestimation of the  $RC$  in the considered section. An interval consisting of an  $RC$  value greater than the unit comprises the fourth grade of the relative risk histogram for  $Q_{mt}$  and characterizes a behavior that is defined, in this study, as physically unacceptable.

## RESULTS AND DISCUSSION

### HHR identification

HHR identification using the geographical convenience method resulted in three groupings, termed 'Geo 1', 'Geo 2' and 'Geo 3', as presented in Figure 3. The grouping 'Geo1' (Figure 3a) was proposed to characterize the physical and statistical behavior of the regionalized flows when the entire area under study was considered to be hydrologically homogeneous, while the groupings 'Geo 2' and 'Geo 3' (Figures 3b and 3c) were proposed because they had the best statistical performances among the groupings tested using the geographical convenience method.

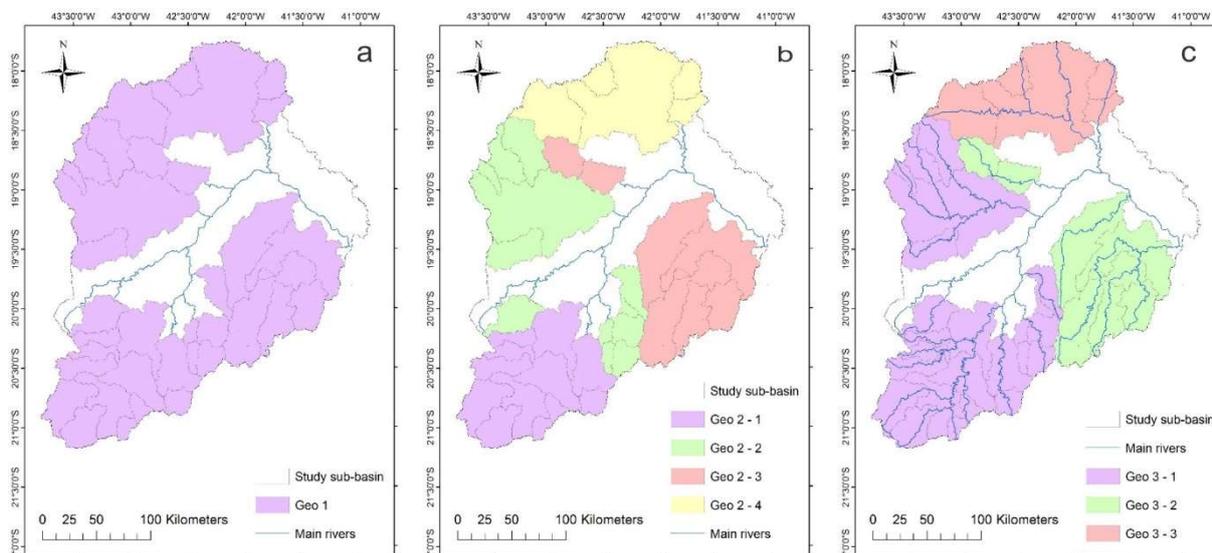


FIGURE 3. HHRs defined on the basis of geographical convenience: 'Geo 1' (a), 'Geo 2' (b) e 'Geo 3' (c).

Regarding the HHR identification using cluster analysis, which is presented in Table 1, the results of the validation indices considering the combinations of variables are as follows: (a) precipitation and evapotranspiration (Cluster 1); precipitation (Cluster 2); latitude, longitude, and real evapotranspiration (Cluster 3); and latitude, longitude and precipitation (Cluster 4). The values in bold

indicate the optimal number of clusters for each validation index.

As listed in Table 1, the validation indices for all of the analyzed groupings are divergent in relation to the recommended number of clusters in which the stations must be divided. Thus, the choice of a number of clusters was conducted considering the prevalence of the four validation indices.

TABLE 1. Comparison of the validation indices for cluster analysis.

	Validation indices	Number of clusters			
		2	3	4	5
Cluster 1	Calinski Harabasz	42.7	38.3	<b>49</b>	46.4
	Dunn	<b>0.16</b>	<b>0.16</b>	0.24	0.25
	Sillhouette	<b>0.45</b>	0.415	<b>0.45</b>	0.43
	Xie_Beni	3.6	2.47	<b>1.42</b>	1.52
Cluster 2	Calinski Harabasz	166	193	215	<b>325</b>
	Dunn	0.41	0.16	0.2	<b>0.15</b>
	Sillhouette	<b>0.72</b>	0.63	0.59	0.62
	Xie_Beni	<b>0.44</b>	2.63	2.56	3.03
Cluster 3	Calinski Harabasz	<b>37.2</b>	30.2	28.9	27.4
	Dunn	0.3	<b>0.19</b>	0.22	0.23
	Sillhouette	<b>0.43</b>	0.33	0.37	0.39
	Xie_Beni	<b>1.44</b>	3.31	1.77	1.79
Cluster 4	Calinski Harabasz	<b>58.6</b>	46.5	46.8	55.6
	Dunn	0.37	<b>0.21</b>	0.3	0.29
	Sillhouette	<b>0.52</b>	0.42	0.44	0.43
	Xie_Beni	<b>0.95</b>	2.45	1.03	1.33

For 'Cluster 1', the validation indices predominantly indicate that the recommended number of clusters is four. For the grouping 'Cluster 2', there is no predominance for the number of clusters because the Silhouette and Xie Beni indices lead to a recommendation of two clusters, while the Calinski Harabasz and Dunn indices indicate a number of clusters equal to five. Therefore, for the grouping 'Cluster 2', both options were considered, as defined as 'Cluster 2/2' and 'Cluster 2/5'.

The validation indices for the groupings 'Cluster 3' and 'Cluster 4' predominantly indicate that the recommended number of clusters for both groupings is two. Therefore, cluster analyses in which the latitude and

longitude variables were included ('Cluster 3' and 'Cluster 4') showed greater agreement in regard to their validation indices compared to analyses in which they were not included ('Cluster 1' and 'Cluster 2').

The groupings 'Cluster 2/2', 'Cluster 3' and 'Cluster 4' have the same configuration in relation to the stream gauge stations, that is, the HHRs formed for these groupings are identical. Thus, the three groupings were represented by the grouping 'Cluster 3'. The grouping 'Cluster 2/5' was renamed to 'Cluster 2'. Briefly, cluster analysis yielded three distinct results, defined as 'Cluster 1', 'Cluster 2' and 'Cluster 3', as presented in Figure 4.

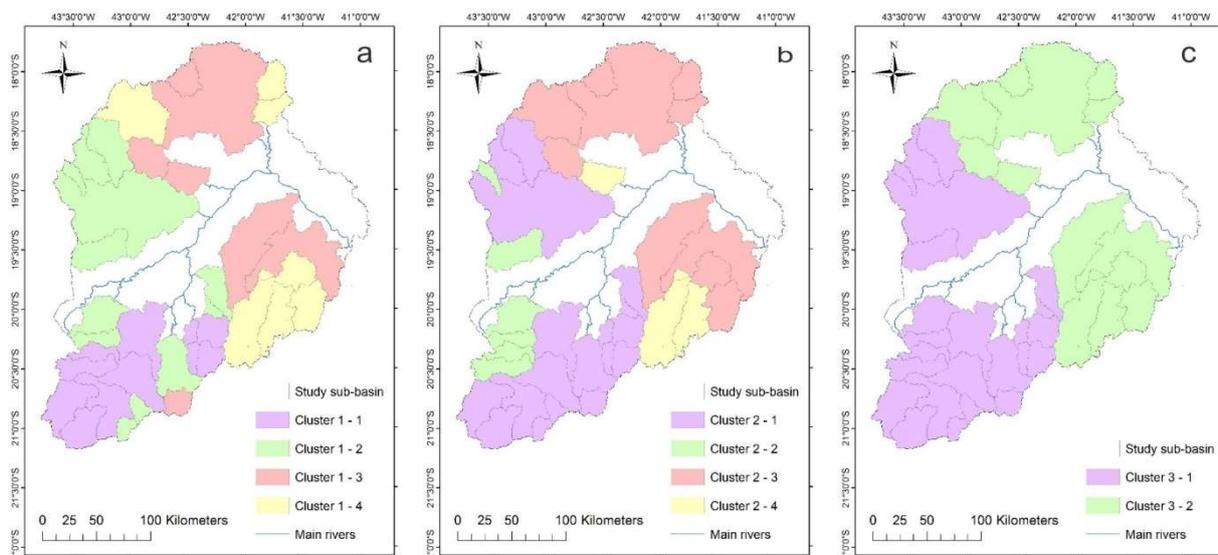


FIGURE 4. HHRs identified on the basis of cluster analysis: 'Cluster 1' (a), 'Cluster 2' (b) e 'Cluster 3' (c).

### Behavior analysis of the regionalized flows

In Table 2, the equations of regression and adjusted coefficients of determination obtained for each HHR identified using the grouping methods in the study are

listed. The methodologies generated regression models with adjusted coefficients of determination higher than 0.85.

The statistical indices used to analyze the behavior of the regionalized flows are listed in Table 3. It was

found that these indices disagreed with showing which grouping method produced a better statistical adjustment. The groupings 'Geo 2' and 'Geo 3' presented the best

statistical performance with regard to  $AE$ ,  $E_1$ ,  $E_2$  and  $ER_m$ . Thus, we considered that these groupings had better statistical results.

TABLE 2. Equations for the average flow regression and coefficients of determination obtained for each HHR defined by the grouping methods.

Grouping	Hydrologically homogeneous region				
	1	2	3	4	5
Geo 1	$0.946 Peq_{750}^{1.002}$ $R^2 = 0.95$	-	-	-	-
Geo 2	$1.161 Peq_{750}^{0.948}$ $R^2 = 0.97$	$0.979 Peq_{750}^{0.993}$ $R^2 = 0.97$	$1.109 Peq_{750}^{0.992}$ $R^2 = 0.91$	$0.439 Peq_{750}^{1.146}$ $R^2 = 0.96$	-
Geo 3	$1.161 Peq_{750}^{0.948}$ $R^2 = 0.97$	$1.109 Peq_{750}^{0.992}$ $R^2 = 0.91$	$0.439 Peq_{750}^{1.146}$ $R^2 = 0.96$	-	-
Cluster 1	$1.190 Peq_{750}^{0.939}$ $R^2 = 0.94$	$0.988 Peq_{750}^{0.995}$ $R^2 = 0.95$	$0.839 Peq_{750}^{1.017}$ $R^2 = 0.98$	$0.176 Peq_{750}^{1.557}$ $R^2 = 0.85$	-
Cluster 2	$1.442 Peq_{750}^{0.890}$ $R^2 = 0.98$	$1.376 Peq_{750}^{0.915}$ $R^2 = 0.95$	$0.913 Peq_{750}^{0.988}$ $R^2 = 0.96$	$0.184 Peq_{750}^{1.557}$ $R^2 = 0.96$	$0.964 Peq_{750}^{0.997}$ $R^2 = 0.87$
Cluster 3	$1.057 Peq_{750}^{0.972}$ $R^2 = 0.97$	$0.592 Peq_{750}^{1.136}$ $R^2 = 0.90$	-	-	-

TABLE 3. Statistical indices obtained for each grouping.

Grouping	$AE$ (%)	$E_1$ (-)	$E_2$ (-)	$E_3$ (-)	$ER_m$ (%)
Geo 1	-32 a 84	0.52	0.33	0.32	18.43
Geo 2	-31 a 31	0.72	0.68	0.41	14.7
Geo 3	-30 a 33	0.71	0.67	0.41	14.7
Cluster 1	-30 a 48	0.66	0.60	0.39	15.85
Cluster 2	-34 a 86	0.67	0.53	0.44	15.48
Cluster 3	-33 a 58	0.56	0.46	0.34	18.6

The group 'Geo 1' showed the worst performance in regards to  $AE$ ,  $E_1$ ,  $E_2$  and  $E_3$  and was considered to be the grouping with the worst statistical result. Taking into account that this grouping corresponds to the basin in the study, which is one hydrologically homogeneous region, this result highlights the importance of identifying HHRs for regionalization studies, mainly for large basins.

Regarding the groupings obtained using cluster analysis, 'Cluster 1' showed a better statistical performance in relation to 'Cluster 3' for all of the evaluated indices, while 'Cluster 2' had a better statistical behavior in relation to 'Cluster 3' in four of the five evaluated indices. Thus, among them, 'Cluster 3' was considered to have the worst statistical result. In regards to 'Cluster 1' and 'Cluster 2', we cannot infer which showed the best statistical performance, as there was no prevalence in the analyzed indices.

### Physical behavior and risk analysis

The spatial distribution of the  $RC$  for the groupings 'Geo 1', 'Cluster 2', 'Cluster 1', 'Geo 2', 'Geo 3' and 'Cluster 3' are presented in Figure 5.

As seen in Figure 5a, it was observed that the grouping 'Geo 1' presented a distribution of  $RC$  consistent with the precipitation map; however, it was found that the extent of the estimated  $RC$  values was from 0.20 to 0.45,

while the  $RC$  values observed in the stream gauge stations ranged from 0.17 to 0.60. Namely, consideration of the area under study as a single HHR led to a reduction in the amplitudes of the estimated  $RC$  values compared to the observed values, causing all of the segments of hydrography to be classified as acceptable/likely in the security analysis. It was also observed that even considering the extrapolation regions of the equation of regionalization, the  $RC$  estimated value amplitudes were much lower than the observed values, which characterizes an inconsistent behavior under a physical point of view. This result is the reason that grouping 'Geo 1' was considered to be unsatisfactory for the regionalization of the  $Q_{mt}$  in the Doce River Basin.

Analysis of the estimated  $RC$  spatial distribution relative to the grouping 'Cluster 2' (Figure 5b) showed that in 17.8% of the hydrography, the  $RC$  values were greater than 1. From a physical point of view, this behavior was also considered to be unsatisfactory for  $Q_{mt}$  regionalization.

$Q_{mt}$  regionalization, considering the grouping 'Cluster 1' (Figure 5c) presented for HHR 4, had estimated  $RC$  values less than 0.01 for 76% of the hydrography, while the smallest  $RC$  value observed in the region was 0.17. This behavior was observed in the hydrography segments within a small drainage area, while for those associated with larger drainage areas, the estimated  $RC$  reached values from 0.31

to 0.45, which are near to those observed in the *RC* stream gauge stations. This behavior is physically inconsistent, as watercourses with estimated *RC* values of approximately 0.01 produce rivers with estimated *RC* values greater than 0.31. A possible physical justification for the occurrence of

this behavior is increased precipitation downstream; however, the precipitation map does not agree with this justification, namely, precipitation decreases downstream. On the basis of this behavior, we considered 'Cluster 1' to not be satisfactory for  $Q_{mlt}$  regionalization.

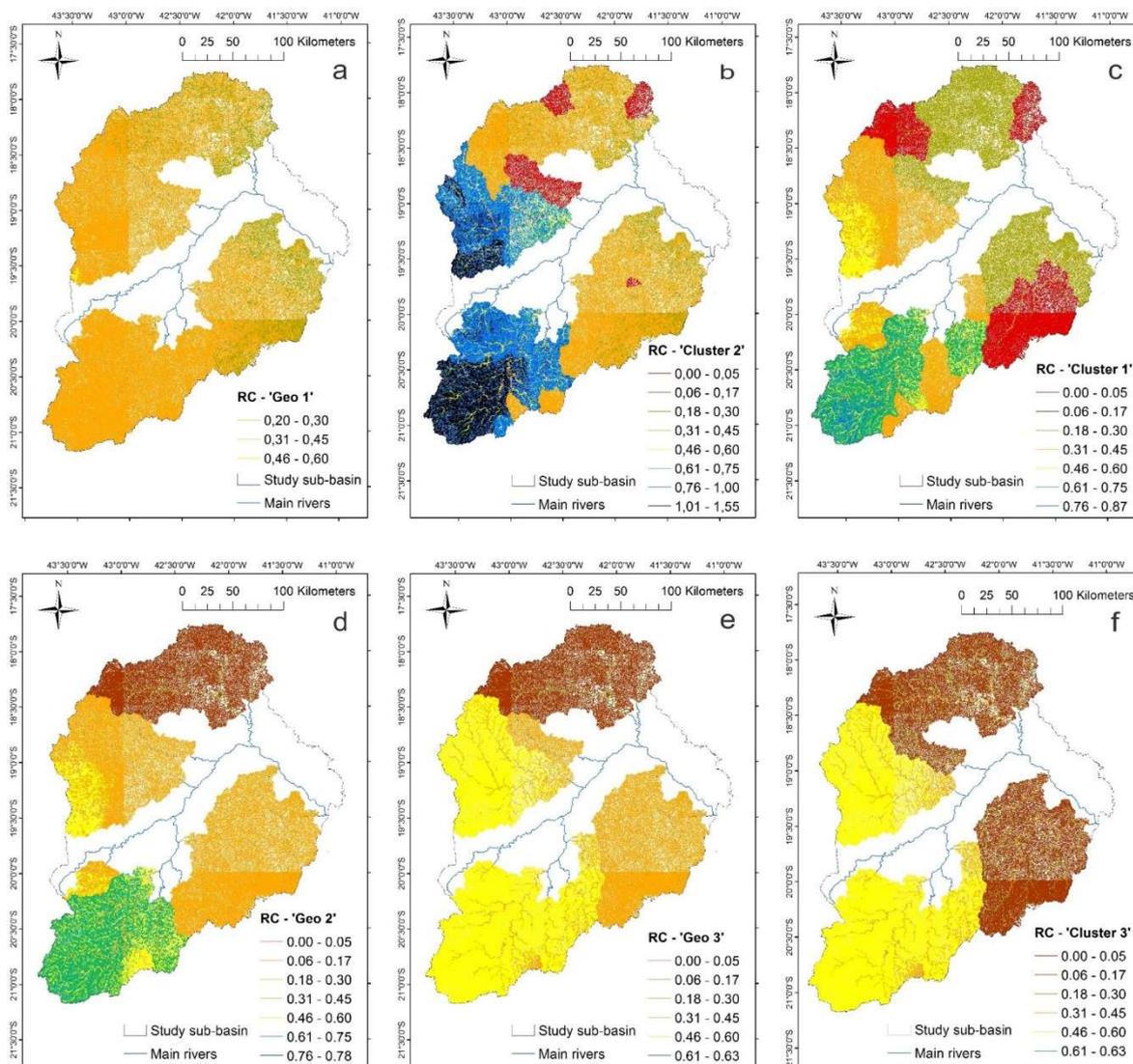


FIGURE 5. Coefficient of spatialized runoff using the equations obtained for the groupings 'Geo 1' (a), 'Cluster 2' (b), 'Cluster 1' (c), 'Geo 2' (d), 'Geo 3' (e) and 'Cluster 3' (f).

From these results, we observe that the physical behavior analysis of regionalized flows assists in determining HHRs and validating regionalized flows because although the groupings 'Cluster 1' and 'Cluster 2' showed better statistical results among the groupings from the cluster analysis (Table 3), their physical behaviors were considered to be unsatisfactory.

The groupings 'Geo 2', 'Geo 3' and 'Cluster 3', unlike the groupings 'Geo 1', 'Cluster 1', 'Cluster 2', were considered to be satisfactory from a physical point of view. However, for some of the hydrography sections, dubious physical behaviors were identified and are described in the following section. This result demonstrates that for the Doce River Basin, acceptable results were found when the HHRs were geographically continuous.

The  $RC$  spatial distribution of the grouping 'Geo 2' (Figure 5d) presented  $RC$  values ranging from 0.03 to 0.78, with an increase in the estimated amplitudes of the  $RC$  value

compared to the interval of variation of the values observed in the stream gauge stations (0.17 to 0.60), which resulted in 23.2% of the hydrography being potentially overestimated and 13.4% of the hydrography being potentially underestimated (Table 4).

The amplitude of the estimated  $RC$  along the hydrography and that observed at the stream gauge stations as well as the average of the estimated  $RC$  values in the hydrography sections and the  $P$  for each HHR of grouping 'Geo 2' are listed in Table 4.

The  $Q_{mlt}$  regionalization from the grouping 'Geo 2' generated the following: (a) 63% of the hydrography sections with  $RC$  values greater than the maximum observed at the stream gauge stations (HHR 1); (b) the average of the estimated  $RC$  values along the hydrography was 0.62, while the maximum value observed in the stream gauge stations was 0.60 (HHR 1); and (c) the minimum estimated  $RC$  value along the hydrography was higher than

the minimum observed at the stream gauge stations (HHR 1 and 2). These behaviors can generate false expectations regarding the availability of water in the respective regions.

The regionalization of  $Q_{mt}$  from the grouping 'Geo 2' presented some inconsistencies between the distribution map and isohyet map. Although the precipitation in HHR 1 and 2 is similar, the difference in the average estimated  $RC$  values along the hydrography is significant (Table 4). A

similar behavior was observed for HHRs 3 and 4. HHR 4 from the grouping 'Geo 2' showed upstream hydrography segments with  $RC$  estimated values smaller than those in the downstream hydrography sections. This behavior is physically dubious, but unlike the grouping 'Cluster 2', the minimum estimated  $RC$  value was 0.03 and average estimated value in the hydrography sections was 0.08.

TABLE 4. Estimated  $RC$  amplitude along the hydrography and observed at the stream gauge stations; the average estimated  $RC$  values in the hydrography section and  $P$  in the HHR grouping 'Geo 2', 'Geo 3' and 'Cluster 3'.

Grouping	HHR	$RC$ amplitude (-)		Average of the estimated $RC$ values (-)	$P$ (mm)
		Observed	Estimated		
Geo 2	1	0.31 a 0.60	0.38 a 0.78	0.62	1317
	2	0.31 a 0.60	0.32 a 0.49	0.43	1288
	3	0.28 a 0.50	0.26 a 0.52	0.38	1125
	4	0.17 a 0.31	0.03 a 0.29	0.08	1165
Geo 3	1	0.31 a 0.6	0.37 a 0.63	0.50	1302
	2	0.28 a 0.5	0.26 a 0.52	0.38	1125
	3	0.17 a 0.31	0.03 a 0.29	0.08	1165
Cluster 3	1	0.31 a 0.6	0.37 a 0.63	0.50	1302
	2	0.17 a 0.5	0.04 a 0.37	0.12	1145

The  $Q_{mt}$  regionalization from the grouping 'Geo 3' (Figure 5e) presented  $RC$  values ranging from 0.03 to 0.63. The percentage of hydrography that was potentially underestimated was 13.17%, and that which was acceptable/likely was 86.63%. HHRs 2 and 3 from 'Geo 3' are identical, respectively, to HHRs 3 and 4 from 'Geo 2'; thus, the considerations are the same as previously reported.

The estimated  $RC$  amplitude along the hydrography and observed at the stream gauge stations, as well as the average of the estimated  $RC$  values of the hydrography sections and  $P$ , in each HHR from grouping 'Geo 3' are listed in Table 4.

The  $Q_{mt}$  regionalization in HHR 1 presented the minimum estimated  $RC$  value along the hydrography as higher than the minimum observed at the stream gauge stations, which leads one to believe that the model overestimates  $Q_{mt}$  at the hydrography sections and thus can generate false expectations of water availability.

The  $Q_{mt}$  regionalization from the grouping 'Cluster 3' (Figure 5f) presented  $RC$  values ranging from 0.04 to 0.63. The percentage of hydrography that was potentially underestimated was 30%, and that within the acceptable/likely zone was 70%. HHR 1 from grouping 'Cluster 3' is identical to that of HHR 1 from grouping 'Geo 3'; thus, the comments are the same as previously reported.

The estimated  $RC$  amplitude along the hydrography and observed at stream gauge stations as well as the average of the estimated  $RC$  values in the hydrography sections and the  $P$  in each HHR from the grouping 'Cluster 3' are listed in Table 4.

HHR 2 from the grouping 'Cluster 3' presented upstream hydrography segments with an estimated  $RC$  smaller than that in the downstream hydrography. This

behavior is physically dubious, but unlike the grouping 'Cluster 2', the minimum estimated  $RC$  value was 0.04.

The regionalization conducted considering the groupings 'Geo 2', 'Geo 3' and 'Cluster 3' was considered to be acceptable from a physical point of view, despite the discussed dubious behaviors. In this way, a comparative analysis between the groupings was conducted to identify which grouping's statistical and physical behaviors presented the lower risk for  $Q_{mt}$  regionalization.

Comparative analysis between the groupings 'Geo 2' and 'Geo 3' showed that the results obtained from the  $Q_{mt}$  regionalization considering the grouping 'Geo 3' to have a better performance in the following respects: a) a greater resemblance between the isohyet map and spatialized  $RC$  behavior (HHR 1); (b) a reduction in the inconsistency noted in Table 4, in which similar precipitation values are associated with distinct estimates for  $RC$ ; and (c) lower estimated  $RC$  values in the southern area of the study, resulting in values that were classified as overestimated being moved to the likely/acceptable zone, reducing the risk associated with regionalization. Regarding the results obtained for the  $Q_{mt}$  regionalization considering the grouping 'Geo 2', we observed that the minimum estimated  $RC$  value in the HHR 2 (0.32) was higher than that observed at the stream gauge stations (0.31), yet less than the estimated value when considering 'Geo 3' (0.37), resulting in, for this region, grouping 'Geo 2' having less risk of overestimating  $Q_{mt}$ . Thus, we could not conclusively determine which of the two groupings represented a lower risk of overestimating the  $Q_{mt}$  regionalization.

A comparative analysis between the groupings 'Geo 2', 'Geo 3' and 'Cluster 3' showed that the results obtained from  $Q_{mt}$  regionalization considering the groupings 'Geo 2' and 'Geo 3' showed a better statistical performance. The

grouping 'Cluster 3' presented one of the worst statistical performances; however, its physical behavior was the most consistent as it showed more similarity between the isohyet map and specialized *RC* behavior, which contributed to a reduction in inconsistencies, as shown in Tables 4, in which similar precipitation values are associated with different *RC* estimates. The comparative analysis between the groupings 'Geo 2', 'Geo 3' and 'Cluster 3' did not highlight which groupings represented less risk for the  $Q_{mt}$  regionalization because while groupings 'Geo 2' and 'Geo 3' showed a superior statistical performance, grouping 'Cluster 3' proved to be more satisfactory when considering the physical behavior of the average regionalized flow.

## CONCLUSIONS

In this study, an analysis of the regionalized long-term average flow physical behavior obtained from HHRs, identified by means of geographical convenience methods and cluster analysis, assisted in choosing the most satisfactory results and in reducing the subjectivity of HHR identification.

The influence of different combinations of latitude, longitude, precipitation and evapotranspiration was examined in the HHRs identified by cluster analysis of the regionalized flows. The combinations that included latitude and longitude presented more compliance for the grouping validation indices, and the physical behavior of the regionalized flows was considered to be satisfactory.

From the results obtained using cluster analysis, the physical analysis was of crucial importance in choosing the HHRs because although 'Cluster 1 and 2' showed better statistical results, their physical behavior was considered to be unsatisfactory. On the other hand, 'Cluster 3' showed the worst statistical performance, but its physical behavior was evaluated as satisfactory.

The HHR delimitation method by geographical convenience produced a better statistical result for regionalized flows; however, the physical result was less than that obtained from cluster analysis.

## ACKNOWLEDGEMENTS

This work was supported by Foundation for Research Support of State of Minas Gerais (FAPEMIG) and the National Council for Scientific and Technological Development (CNPq).

## REFERENCES

- Agarwal A, Maheswaran R, Sehgal V, Khosa R, Sivakumar B, Bernhofer C (2016) Hydrologic regionalization using wavelet-based multiscale entropy method. *Journal of Hydrology* 538:22-32.
- Allen RG, Pereira LS, Raes D, Smith M (1998) Crop evapotranspiration: guidelines for computing crop water requirements. Rome, FAO. Irrigation and drainage paper 56.
- Arsenault R, Brissette FP (2014) Continuous streamflow prediction in ungauged basins: The effects of equifinality and parameter set selection on uncertainty in regionalization approaches. *Water Resources Research* 50(7): 6135-6153.
- Beskow S, de Mello CR, Faria LC, Simões MC, Caldeira TL, Nunes GS (2014) Índices de sazonalidade para regionalização hidrológica de vazões de estiagem no Rio Grande do Sul. *Revista Brasileira de Engenharia Agrícola e Ambiental-Agriambi* 18(7):748-754.
- Calinski T, Harabasz J (1974) A dendrite method for cluster analysis. *Communications in Statistics*. 3(1):1-27.
- Coxon G, Freer J, Westerberg IK, Wagener T, Woods R, Smith PJ (2015) A novel framework for discharge uncertainty quantification applied to 500 UK gauging stations. *Water resources research* 51(7):5531-5546
- Dunn J (1974) Well separated clusters and optimal fuzzy partitions. *Journal of Cybernetics* 4(1):95-104.
- Gonçalves CJ, Oliveira ACM de, Oliveira JRS de, Ribeiro RB (2018) Estudo de regionalização de vazões para a bacia hidrográfica do rio Paranaíba. *Sustentare* 2(2):97-114.
- GPRH - Grupo de Pesquisas em Recursos Hídricos, IGAM - Instituto Mineiro das Águas (2012) Estudo de regionalização de vazões para o aprimoramento do processo de outorga no Estado de Minas Gerais. GPRH, p252-299.
- Hosking JRM, Wallis JR (1997) Regional frequency analysis – an approach based on L-moments. Cambridge, Cambridge University Press, 224p.
- Kim D, Jung I, Chun JA (2016) A comparison between parameter regionalization and model calibration with flow duration curves for prediction in ungauged catchments. *Hydrology and Earth System Sciences Discussions* 5(8):1-29
- Krause P, Boyle DP, Base F (2005) Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences* 5:89-97.
- Kult JM, Fry LM, Gronewold AD, Choi W (2014) Regionalization of hydrologic response in the Great Lakes basin: Considerations of temporal scales of analysis. *Journal of Hydrology* 519:2224-2237.
- Liew MWV, Mittelstet AR (2018) Comparison of three regionalization techniques for predicting streamflow in ungauged watersheds in Nebraska, USA using SWAT mode. *International Journal of Agricultural and Biological Engineering* 11(3):110-119.
- Louzada FLRO, Xavier AC, Pezzopane JEM (2018) Climatological water balance with data estimated by tropical rainfall measuring mission for the Doce river basin. *Engenharia Agrícola* 38(3):376-386.
- Naghttini M, Pinto JEA (2007) Hidrologia estatística. Belo Horizonte, CPRM, 552p.
- Nathan RJ, McMahon TA (1990) Identification of homogeneous regions for the purposes of regionalization. *Journal of Hydrology* 121(1-4):217-238.
- Novaes LF (2005) Modelo para a quantificação da disponibilidade hídrica na Bacia Hidrográfica do Rio Paracatu. Msc Thesis, Federal University of Viçosa.

- Nruthya K, Srinivas VV (2015) Evaluating methods to predict streamflow at ungauged sites using regional flow duration curves: A case study. *Aquatic Procedia* 4:641-648.
- Passos MLV, Zambrzycki GC, Pereira RS (2017) Balanço hídrico climatológico e classificação climática para o município de Balsas-MA. *Scientia agraria* 18(1):83-89.
- Pruski FF, de Araújo AN, Pruski PL, Rodriguez RDG (2013) Improved regionalization of streamflow by use of the streamflow equivalent of precipitation as an explanatory variable. *Journal of Hydrology* 476:52-71.
- Pruski FF, Rodrigues RG, Pruski PL, Nunes AA, Rego FS (2016) Extrapolation of regionalization equations for long-term average flow. *Engenharia Agrícola* 36(5):830-838.
- Pruski FF, Rodriguez RDG, Nunes AA, Pruski PL; Singh VP (2015) Low-flow estimates in regions of extrapolation of the regionalization equations: a new concept. *Engenharia Agrícola* 35(5):808-816.
- Rao AR, Srinivas VV (2006) Regionalization of watersheds by hybrid-cluster analysis. *Journal of Hydrology* 318(1):37-56.
- Razavi T, Coulibaly P (2013) Streamflow prediction in ungauged basins: review of regionalization methods. *Journal of Hydrologic Engineering* 18(8):958-975.
- Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20:53-65.
- Samuel J, Coulibaly P, Metcalfe RA (2011) Estimation of Continuous Streamflow in Ontario Ungauged Basins: Comparison of Regionalization Methods. *Journal of Hydrologic Engineering* 16(5):447-459.
- Şen Z, Altunkaynak A (2006) A comparative fuzzy logic approach to runoff coefficient and runoff estimation. *Hydrological Processes: An International Journal* 20(9):1993-2009.
- Silva Júnior OB, Bueno EDO, Tucci CE, Castro NM (2003) Extrapolação espacial na regionalização de vazão. *Revista Brasileira de Recursos Hídricos* 8(1):21-37.
- Smakhtin VU (2001) Low flow hydrology: a review. *Journal of Hydrology* 240(3-4):147-186.
- Thornthwaite CW, Mather JR (1955) *The water balance*. Centerton, NJ: Drexel Institute of Technology, Laboratory of Climatology, 104p.
- Westerberg IK, Gong L, Beven KJ, Seibert J, Semedo A, Xu CY, Halldin S (2014) Regional water balance modelling using flow-duration curves with observational uncertainties. *Hydrology and Earth System Sciences* 18(8):2993-3013.
- Xie XL, Beni G (1991) A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13(8):841-847.