Streamflow In The Sapucaí River Watershed, Brazil: Probabilistic Modeling, Reference Streamflow, And Regionalization

Marcel Carvalho Abreu [1], Micael de Souza Fraga [2], Laura Thebit Almeida [3], Felipe Bernardes da Silva [4], Roberto Avelino Cecílio [5], Gustavo Bastos Lyra [1] and Rafael Coll Delgado [1]

[1] Federal Rural University of Rio de Janeiro, Forest Institute, Department of Environmental Sciences, Seropédica, Rio de Janeiro, Brazil.

[2] Water Management Institute of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

[3] University of Viçosa, School of Agriculture, Agricultural Engineering Department, Viçosa, Minas Gerais, Brazil.

[4] Vale do Rio Verde University, Três Corações, Minas Gerais, Brazil

[5] Federal University of Espirito Santo, Department of Forest and Wood Sciences, Jerônimo Monteiro, Espírito Santo, Brazil.

Corresponding Author: Marcel Carvalho Abreu - Rod. BR 465, Km 07 Seropédica - RJ - CEP: 23890-000 Universidade Federal Rural do Rio de Janeiro - UFRRJ Instituto de Florestas - IF/Departamento de Ciências Ambientais – DCA. E-mail: marcelc.abreu@gmail.com

Abstract

This work aims to study the streamflow statistic patterns in the Sapucaí River watershed, state of Minas Gerais, Brazil. This study embraces the streamflow probabilistic modeling to determine the reference streamflow and, later, the streamflow regionalization to improve the water resources management. A 26-year-data series (1989 - 2014) of maximum, average, and minimum streamflow were used. Probability density functions were applied to the maximum and minimum daily streamflow to determine the recurrence periods. Long-term average annual and monthly streamflow were also calculated. Linear and non-linear regressions were adjusted for the streamflow regionalization. The drainage area and the streamflow equivalent to the total rainfall (with and without abstractions) were used as predictor variables. The probability density functions that best adjusted the maximum streamflow data set were the Generalized

Extreme Values, and for the minimum streamflow was the normal distribution. Linear and non-linear regressions were efficient (R^2> 0.90 and d Willmott> 0.97) in the regionalization process regardless of the predictor variables. However, a small statistical advantage was found for the adjustment of non-linear regressions that used the predictor variables drainage area and the streamflow equivalent to the total rainfall (without abstractions).

Keywords: Reference water flow, Streamflow probabilistic modeling, Hydrology, 10.

1. INTRODUCTION

Water is an extremely important element human activity. The water availability has been reducing over the years due to the increase in population density, expansion of irrigated agriculture, and degradation of the water quality (Pruski, Nunes, Pruski, & Rodriguez, 2013). In addition, larger regional-scale trends in floods (Mediero et al., 2015) and water availability (Koutroulis et al., 2019) often result from changes in climatic variables. Governments and international agencies highlight water as the most important natural resource (Silva, Oliveira, Mello, & Pierangeli, 2006) and point out the need for its better management of the water resources. Adequate management requires, fundamentally, the availability of hydrological data, from a dense hydrometric network uniformly distributed in space and with consistent hydrological data (Beskow et al., 2014). Currently, these requirements are not met in Brazil (Baena, Silva, Pruski, & Calijuri, 2004; Piol, Reis, Caiado, & Mendonça, 2019; Pruski, Rodriguez, & Nunes, 2015). Therefore, some statistical/hydrological techniques have been developed or adapted to provide estimations for hydrological data in places where measured data is scarce. Such techniques are hydrological models (Andrade, Mello, & Bescow, 2013; Beskow, Norton, & Mello, 2013) and hydrological regionalization (Baena et al. 2004; Maciel, Vieira, Monte, & Vasques, 2019; Pruski et al., 2015; Pruski, Rodriguez, Pruski, Nunes, & Rego, 2016).

Hydrological regionalization is an useful technique to compensate the lack of hydrological data in places where this data is scarce or non-existent (Beskow et al., 2014; Pruski et al., 2013). This is done to support water resources management (Piol et al., 2019). The streamflow stands out among the hydrological data important for the water resources management (Barros, Pessoa, Santana, Lopes, & Costa, 2018; Cecilio, Zanetti, Gasparini, & Catrinck, 2018; Costa et al. 2019; Pruski et al., 2015). Although streamflow data is essential for the water use grant rights process, most Brazilian states still lack detailed streamflow data (Lisboa, David, Moreira, Silva, & Uliana, 2019).

The long-term average streamflow is an important variable used to characterize the water potential to regularize the streamflow. Its determination needs a historical series with a considerable number of years (at least 20 years), which is not always available (Pruski et al., 2013). The minimum streamflow is used to determine the water availability and its values serve as a reference to set limits for water grants. In Minas Gerais (Lisboa et al., 2019), and São Paulo (Wolff, Duarte, & Mingoti, 2014) states, southeastern Brazil, the minimum reference streamflow used to limit the water grant is based on the average minimum streamflow for seven consecutive days with a 10-year recurrence period ($Q_{7,10}$). To determine $Q_{7,10}$, beyond the availability of a long and consistent historical data series, a probability density function (PDF) is also needed to obtain the theoretical frequency associated with the 10-year recurrence period. Therefore, it is necessary to verify different PDFs and their performances in representing the minimum streamflow data set by using goodness-of-fit tests (Barros et al., 2018; Finkler, Mendes, Schneider, Bortolin, & Schneider, 2015).

In Brazil, at a federal level, the minimum reference streamflow is the one that remains in the watercourse for, at least, 95% of the time (Q_{95}) (Serrano, Ribeiro, Borges, & Pruski, 2020). $Q_{7,10}$ and Q_{95} are widely used to set limits for water withdrawals. Even though they represent different conditions in terms of water limitation, the $Q_{7,10}$ represents more extreme conditions for minimum streamflow than Q_{95}, both are important tools for water resources management and planning (Ouyang, 2012; Serrano et al., 2020). Other reference minimum streamflow that is important to the water resources management and planning is the streamflow that remains in the watercourse for 90% (Q_{90}), 80% (Q_{80}), and 50% (Q_{50}) of the time (Baena et al., 2004). Despite advances in statistical and process−based hydrological models, The estimation of low flows in rivers is a vexing problem (Konrad & Rumsey, 2019).

The maximum streamflow is of great importance in the design of hydraulic projects and flood predictions (Lopes, Prado, Zolim, Paulino, & Antoniel, 2016; Mediero et al., 2015). Usually, the maximum streamflow is associated with a recurrence period that indicates the project's safety. The higher the project's safety the higher the recurrence period. However, the higher the recurrence period the more expensive is the construction (Cassalho et al., 2017). The recurrence periods usually used for hydraulic projects range between 5 and 500 years depending on the project.

For the hydrological regionalization of streamflow, several methods are used, such as the traditional method (ELETROBAS 1985), the methods of linear interpolation and modified linear interpolation (ELETROBRAS 1985), the characteristics values method, and the exponential curve method (Piol et al., 2019). Among these methods, the traditional method stands out. It starts with the identification of hydrologically homogeneous regions. Later, linear or non-linear regressions are applied using the morphometric and/or climatic characteristics (predictor variables) and the targeted streamflow (response variable). Linear (multiple or simple) or non-linear (power or exponential) regressions using drainage area (Da), and/or the streamflow equivalent to the total rainfall (P_{eq}), or the streamflow equivalent to the rainfall volume considering the abstraction of part of the rainfall that does not reach the river and does not become streamflow (P_{eq750}) are the most used ones (Cassalho et al., 2017; Pruski et al., 2015). The traditional method has been successfully applied and has superior performance than the other methods in some watersheds located in several regions of Brazil e.g. (Cecilio et al., 2018; Amorim et al., 2020; Matos, Uliana, Martins, & Rapalo, 2020).

The Sapucaí River watershed is part of the Grande River watershed, located in southeastern Brazil, crossing São Paulo and Minas Gerais states. The Sapucaí River watershed is in an important Brazilian region, with a predominance of the sectors of services, industry, and agriculture. The region's GDP between 2016 and 2017 corresponded to near 1.2% of Brazil's GDP in the same period. The region has a complex topography with an orographic influence of the Serra da Mantiqueira, which is crucial for the rainfall regime, the river formation, and the hydrological regime. In general, the water resources in this watershed are not intensely used (low hydrological stress) (Duraes, Mello, & Bescow, 2015), however, the watershed has flood-prone regions (Almeida, Abreu, Fraga, Silva, & Cecílio, 2017). The Sapucaí River watershed's mouth is the Furnas dam, a hydroelectric power plant with an area of 1,440 km², 1,216 megawatts of power, and intersecting 34 cities in Minas Gerais, one of the main power plants in Brazil.

Although the Sapucaí River watershed has interesting hydrological conditions regarding water resources availability and flood-prone areas, few studies focused on its average, minimum, and maximum streamflow, which is the motivation of this study. In light of the aforementioned, this work aimed to study the streamflow in the Sapucaí River watershed, with the specific goals: i) determine the long-term average streamflow, the maximum annual daily streamflow and the minimum reference streamflow ($Q_{7,10}$, Q_{95}, Q_{90}, Q_{80}, Q_{50}); ii) test different probability density functions in the representation of the maximum and minimum reference streamflow data using goodness-of-fit tests; and iii) obtain models for regionalization of the long-term average, maximum, and minimum reference streamflow.

2. MATERIALS AND METHODS

2.1 Study area

The Sapucaí River watershed is located in the southeastern region of Brazil (Figure 1), with a drainage area of 25,095.79 km² (Almeida et al., 2017). It covers part of the states of São Paulo, where the Sapucaí River begins in Campos do Jordão, state of São Paulo. It also covers part of Minas Gerais, where it ends in the Furnas dam (Almeida et al., 2017; Matos, Pioltine, Mauad, & Barbosa, 2011). The Furnas dam power plant can produce 1,216 Mega Watts of energy and is of great importance for the Brazilian energy scenario (Durães & Mello 2016). The watershed includes more than 75 municipalities and a population of 1,615,128 inhabitants (IBGE 2019). The region's economy is concentrated in the provision of services, industries and agriculture, the latter activity being extremely relevant to the socioeconomic dynamics of the region. The production of coffee, dairy farming, metallurgy-aluminum, mining, agribusiness, electronics, helicopters, auto parts, beverages, textiles, and tourism stands out (IBGE 2010; 2017).

The predominant soils are the Inceptisols (≈56%), Oxisols (≈32%), Entisols (≈11%) and Ultisols (≈1%). The land cover/land use include mostly pastures, area of native vegetation, mostly Atlantic Forest, agriculture, planted forests (especially Eucalyptus spp forets) and urban area (Durães & Mello 2016).

The region's climate is classified, according to Köppen, as subtropical with dry winter and hot summer (Cwa) and subtropical of altitude with dry winter and temperate summer (Cwb) (Alvares, Stape, Sentelhas, Gonçalves, & Sparovek, 2013; Martins, Gonzaga, Santos, & Reboita, 2018). The region has great orographic influence due to Serra da Mantiqueira, with elevations ranging from 774 to 2795 m, average slope of 16% (wavy relief) with the slope in some areas ranging from 0% (flat relief) to 218% (strongly mountainous relief) (Almeida et al., 2017). Annual total rainfall varies between 1,500 and 1,700 mm, and the average annual air temperatures between 15 and 19 ºC.

2.2 Data acquiring, selection and pre-treatment

Fourteen streamflow gauges and five rain gauges (Figure 1 and Table 1) were selected from the National Water and Basic Sanitation Agency (ANA), available on the hydrological data platform HidroWeb (Hydrological Information System - http://www.snirh.gov.br/hidroweb/), and the National Institute of Meteorology (https://portal.inmet.gov.br/), respectively. The gauges were within the Sapucaí River watershed and had consistent daily data. The base period, from 1989 to 2014, met at least 20-year-data minimum criteria (Pruski et al., 2016, 2015).

[Insert Table 1]

[Insert Figure 1]

Afterward, series of maximum annual daily streamflow (Q_{max}), average annual streamflow (Q_{avg}), average monthly streamflow (Q_{Jan}, Q_{Feb}, …, Q_{Dec}), and minimum streamflow averaged from seven consecutive days (Q_7) were constructed for each stream gauge station. Minimum streamflow of 95% (Q_{95}), 90% (Q_{90}), 80% (Q_{80}), and 50% (Q_{50}), from the permanence curve, were also established. The permanence curve is a hydrological function that relates flow rate and the percentage of time that this flow is equal or exceeded during the entire historical period considered for its construction. The gauge selection and the annual series were aided by the Computational System for Hydrological Analysis (SisCAH 1.0), developed by the Research Group on Water Resources at the Federal University of Viçosa (Sousa, Pruski, Bof, Cecon, & Sousa, 2009).

The rainfall data analysis consisted in obtaining the watershed's average annual rainfall (P, mm) in a way that P represented the rainfall in the entire watershed. Later the streamflow equivalent to the total rainfall (P_{eq}) was calculated (Pruski et al., 2013, 2015, 2016). It was also calculated the streamflow equivalent to the rainfall volume considering the abstraction of part of the rainfall (750 mm) that does not reach the river and does not become streamflow (P_{eq750}) (Pruski et al., 2016). These variables are commonly used as independent variables in regionalization studies and are obtained by applying the equations 1 and 2:

$$P_{eq}=\frac{P\cdot \mathrm{Da}}{k} \qquad (1)$$

$$P_{eq750}=\frac{(\mathrm{P\text{-}750})\cdot \mathrm{Da}}{k} \qquad (2)$$

In which: P_{eq} is the equivalent streamflow for the average annual rainfall (m3 s-1); P_{eq750} is the equivalent streamflow for the average annual rainfall considering the abstraction of 750 mm of the rainfall ($m^3\ s^{-1}$); P is the watershed's average annual rainfall (mm); Da is the drainage area upstream of the cross-section of interest (km2); and k is a conversion factor equal to 31,536.

2.3 Statistical Analyses

Statistics of position and dispersion were calculated for the series of Q_{max}, Q_{avg}, Q_7, and average monthly streamflow (Q_{Jan}, Q_{Feb}, ... Q_{Dec}). The box-plot analysis was used to check extreme values, quartiles, Q_{max} and Q_7 patterns in different gauges, as well as their mean, standard deviations, kurtosis coefficients, and asymmetry coefficient. The average streamflow was analyzed based on mean, standard deviation, kurtosis coefficient (k) and asymmetry (g), which was calculated according to equations 3 and 4, respectively (Naghettini & Pinto 2007):

$$k = \frac{N^2}{(N-1)\cdot(N-2)\cdot(N-3)} \cdot \frac{\sum_{i=1}^{N} (x_i - \overline{x})^4}{s^4}$$

(3)

$$g = \frac{N}{(N-1)\cdot(N-2)} \cdot \frac{\sum_{i=1}^{N} (x_i - \overline{x})^3}{s^3}$$

(4)

In which: xi are the observed streamflow, $\overline{X}$ is the streamflow's mean, s is the standard deviation, and N is the number of observations.

Probability density functions (PDFs) were applied to the series of Q_{max} and Q_7 to associate the streamflow with a frequency of occurrence represented by the recurrence period (RT). The PDFs were applied according to Naghettini e Pinto (2007) and are shown in Table 2. The recurrence periods of 5 (Q_{max5}), 10 (Q_{max10}), 20 (Q_{max20}), 50 (Q_{max50}), and 100 (Q_{max100}) years associated with the maximum annual streamflow. The 10-year recurrence period ($Q_{7,10}$) associated with the minimum streamflow of time permanence in the watercourse.

Goodness-of-fit tests were used to verify the adherence of the PDFs to the streamflow data. The null hypothesis (H_0) states that the probabilistic pattern of the random variable can be modeled by the tested probability function. The alternative (H_1) states that the probabilistic pattern of the random variable cannot be modeled by the probability distribution function tested.

[Insert Table 2]

To test the mentioned hypotheses the following adherence tests were used: Kolmogorov-Smirnov (KS), chi-square (χ^2), Anderson-Darling (AD), Cramer-Von Mises (CVM), Filliben (Fi) and Shapiro-Wilk (SW). The goal of using several adhesion tests was to verify the rigor of these tests and the PDFs versatility in representing the maximum and minimum streamflow data, in several adhesion tests (regardless of the test and its rigor). The reason for it is because each goodness-of-fit test has its particularities.

Table 3 shows each test statistics to be compared with the critical standard value associated with a significance level. In this study, the significance level adopted was 5% according to several hydrological studies (Abreu et al., 2018; Beskow et al., 2015; Granemann, Hofherr, & Merz, 2018; Costa et al., 2019). The probabilistic modeling of maximum and minimum streamflow was performed using the "EnvStats" package (Millard, 2013) from the R software (R Team Core, 2018).

[Insert Table 3]

The analysis of aggroupment (cluster) and the analysis of the reference streamflow normalized according to the drainage area were the statistical procedures used to verify homogeneous regions in terms of streamflow. Cluster analysis was used to group hydrologically homogeneous regions in terms of streamflow in several studies (Hannaford, Buys, Stahl, & Tallaksen, 2013; Elesbon et al., 2015; Mediero et al., 2015). Ward's hierarchical method was one of the most used methods (Mediero et al., 2015). The number of established groups was determined by a model based on the parameterized finite Gaussian mixture, in which the models are estimated by the algorithm of maximum optimization expectation (iterative method to find parameter estimates with maximum likelihood) that base the clustering in a hierarchical model. The ideal model is then selected according to the Bayesian Information Criterion (BIC). The model with the lowest BIC is the one with the best adjustment. The cluster analysis was performed with the R software through the "mcluster" package (Scrucca, Fop, Murphy, & Raftery, 2016). The main reference streamflow used in the cluster analysis of this study were Q_{avg}, Q_{max}, Q_7, Q_{90}, and Q_{95}.

The analysis of each gauge normalized streamflow was also used as a parameter to verify the hydrologically homogeneous regions. To this end, the reference streamflow was plotted as a function of the drainage area. A similar pattern is expected in the same homogeneous region between streamflow and drainage area (same angular coefficient) (ELETROBRAS 1985).

The traditional method is widely used for streamflow regionalization (Pruski et al., 2015, 2016). It consists of relating the streamflow (Q) with the watershed's characteristics, such as the gauge's drainage area (Da), drainage density (Dd), slope (S), river length (C), and/or climatic characteristics, such as rainfall (P), streamflow equivalent to the total rainfall (Peq) or the streamflow equivalent to the rainfall volume considering the abstraction of part of the rainfall that does not reach the river and does not become streamflow (Baena et al., 2004; Pruski et al., 2015):

$$Q = f(Da, P_{eq}, P_{eq750}) \tag{5}$$

Since the variables Da, P_{eq}, and P_{eq750} are easily obtained and have good performance as a predictive variable (ordinate axis) of streamflow (abscissa axis) (Cecilio et al., 2018), they were used for the regression analysis. The regression models tested were first-order linear regression and nonlinear power regression. The quality of the adjustment of the regressions established for the streamflow regionalization was made through the determination coefficient (R^2), the Willmott concordance index (d) and the root means square error (RMSE).

$$R^2 = \left[\frac{\sum_{i=1}^{N} \left(O_i - \overline{O}\right) \cdot \left(E_i - \overline{E}\right)}{\sqrt{\sum_{i=1}^{N} \left(O_i - \overline{O}\right)^2 \cdot \sum_{i=1}^{N} \left(E_i - \overline{E}\right)}} \right]^2 \tag{6}$$

$$d = 1 - \left[\frac{\sum_{i=1}^{N} \left(O_i - E_i\right)^2}{\sum_{i=1}^{N} \left(\left|E_i - \overline{O}\right| + \left|O_i - \overline{O}\right|\right)^2} \right] \tag{7}$$

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{\left(E_i - O_i\right)^2}{n}} \tag{8}$$

3. RESULTS

3.1 Streamflow position and dispersion statistics

The monthly and annual long-term average streamflow for the 26 years analyzed are shown in Table 4. Table 4 is of great importance for the development of water use projects (e.g. water supply, energy production, irrigation, navigation) because it represents the potential water availability in the watershed (Pruski et al., 2016). In general, the highest streamflow is observed along the Verde River (gauges: 61537000, 61510000 and 61460000), and the Sapucaí River (61305000) in Santa Rita do Sapucaí, before joining with the Sapucaí Mirim River (Figure 1). The two lowest streamflow were found in the tributaries of the Sapucaí Mirim River (gauges: 61250000 and 61295000), in the upper region of the Sapucaí River and just after the Sapucaí Mirim joins the Sapucaí River.

[Insert Table 4]

Table 4 analysis allows identifying the effects of the streamflow seasonality influenced by the rainy season. Usually, the streamflow decreases from March until reaching its minimum values in August then it starts to increase until the months of greatest streamflow, usually January and February. The kurtosis coefficient had positive values for all months, and its values were close ($3.2 \leq k \leq 5.1$). The asymmetry coefficient followed the same pattern with positive values and little variation ($1.9 \leq g \leq 2.2$). Therefore, the monthly series and the long-term averages (Q_{avg}) are considered leptokurtic (its distribution function curve is larger than the normal distribution) and with its tail longer on the right side than the left side (greater number of streamflow observations in the smallest classes).

Figure 3 shows the box-plot analysis of the Q_{max} (Figure 3a), and the Q_7 (Figure 3b). The highest Q_{max} and Q_7 were found in the Verde River (61537000 and 61510000) in the Três Corações, and Careaçu region, and in the Sapucaí River (61305000) in Santa Rita do Sapucaí. The lowest Q_{max} (61250000, 61295000 and 61343000), and Q_7 (61295000, 61343000, and 615650000) were found in gauges at the south of the watershed, in tributaries of the Sapucaí River or the Verde River (Figure 1 and Table 1).

[Insert Figure 3]

Table 5 shows the kurtosis (k) and asymmetry (g) coefficients for maximum streamflow, and minimum streamflow of seven consecutive days for the gauges in the Sapucaí River watershed. In general, the Q_{max} distributions showed, mostly, a leptokurtic pattern on the right. It means that the Q_{max} distributions are fewer and has most of the observations in smaller streamflow classes. The Q_7 showed both platykurtic series (more tapered than the normal distribution) on the right (concentrates greater streamflow observations in smaller classes), and leptokurtic series (fewer than the normal distribution) on the left (concentrates greater streamflow observations in higher streamflow classes), and the right.

[Insert Table 5]

3.2 Streamflow probabilistic modeling

Table 6 shows each PDF percentage of adherence in each goodness-of-fit test. The goodness-of-fit tests showed differences in rigor regarding the acceptance of the null adherence hypothesis. However, the same pattern regarding each PDF's suitability to represent the minimum and maximum streamflow in each gauge were observed. The KS test was the most permissive in accepting adherence, followed by the χ^2 test. The CVM and AD tests showed intermediate rigor, while the Fi and SW tests were, respectively, the most rigorous ones. It is important to state that the PDF GEV couldn't be adjusted to represent the minimum streamflow from the Itajubá gauge (61272000), through the parameter estimation using the maximum likelihood method. Lyra et al. (2006) also observed that the KS test had a lower tolerance than the χ^2 for assessing the PDF adherence to the monthly rainfall. The authors argue that this is because KS compares only the most frequent occurrence classes, while χ^2 compares all classes.

[Insert Table 6]

Analysing the PDF's performance in representing the Q_{max} and Q_7 data, a small difference was found between the distributions. For the maximum streamflow the distributions that stood out, in performance order, were the GEV, Gamma, and Weibull (99%, 86% and 83% adherence, respectively) and for Q_7 stood out the distributions Normal, Weibull and GEV (98%, 95%, and 94% adherence, respectively).

Figure 4 shows the p-values classification obtained in the different goodness-of-fit tests to represent the maximum streamflow. It is noted that the highest p-values were found in the Kolmogorov-Smirnov test (Figure 4a) which is less rigorous, while the Filliben (Figure 4e) and Shapiro-Wilk tests (Figure 4f) were the most rigorous ones. The other tests (χ^2, AD e CVM) had intermediate rigor. The GEV distribution had the highest p-values regardless of the goodness-of-fit tests. It confirms that this PDF had the best performance in representing the streamflow data, especially maximum streamflow. The Gamma and Weibull distributions also had good results in terms of adherence to the Q_{max} data. The normal distribution had the worst performance in representing the Q_{max} data.

[Insert Figure 4]

The p-values found in the PDFs representing Q_7 are shown in Figure 5. The pattern of the rigor of the goodness-of-fit tests was similar to the one obtained for Q_{max}. The highest p-values were found in the KS test. (Figure 5a) and the lowest in the rFi tests (Figure 5e) and SW (Figure 5f). However, regarding the probability distributions, performance patterns found in the Q_{max} goodness-of-fit tests were different. The Normal distribution had the best performance in representing the minimum streamflow, followed by the GEV and Weibull distribution. The Log-normal and Gumbel distributions had the worst performances. Gamma PDF had an intermediate performance. Comparatively, the Q7 had the coefficients k and g values closer to zero than the Q_{max} which indicates greater proximity to the normal distribution.

[Insert Figure 5]

Tables A1 and A2 show the parameters of the probability density functions (PDF) for the maximum and minimum reference streamflow, respectively. For Q_{max}, the parameters of location and scale of GEV distribution were high correlation (r) with Da (r = 0.998 and r = 0.987), P_{eq} (r = 0.998 and r = 0.988) and P_{eq750} (r = 0.998 and r = 0.987) indicates the position of the peak (class of the peak) and the magnitude of the peak. The correlations between Ad, P_{eq}, P_{eq750} Q_{95}, Q_{90}, Q_{80} and Q_{95} with the mean and standard derivation parameters of normal distributions were high (r > 0.97). It represents the degree to which the minimum reference flows are dispersed around the average.

According to the results obtained in this study, the maximum streamflow representation was made by the GEV distribution, while the minimum streamflow representation was done by the normal distribution.

3.3 Streamflow probabilistic modeling

The analysis of homogeneous regions in terms of reference streamflow through cluster analysis and the analysis of normalized frequencies showed that it is not necessary to discriminate the Sapucaí River watershed in groups. Figure 6a shows the performance of the models in predicting the number of clusters. The ideal model was the ellipsoidal, equal volume, shape, and orientation (EEE), with the lowest BIC index in a single group. Figure 6b shows the pattern of the maximum, average, and minimum streamflow in function of the drainage area. It is notable the linearity of the streamflow in function of the drainage area. Thus, regionalization through the traditional method can include all gauges selected in this study.

Table 7 shows the linear and non-linear (power) regression models for the regionalization of maximum streamflow for the recurrence periods of 5 (Q_{max5}), 10 (Q_{max10}), 20 (Q_{max20}), 50 (Q_{max50}) and 100 (Q_{max100}) years. In general, the models were able to predict the maximum streamflow with precision (R^2> 0.90) and accuracy (d> 0.97). The errors represented by the RMSE were lower than 25, 40, 56, 81, and 103 m^3 s^{-1}, for the recurrence periods of 5, 10, 20, 50, and 100 years, respectively. The performance of streamflow regionalization models was similar to that one of other studies with the same goals, which proves their feasibility (Baena et al., 2004; Lopes et al., 2016; Cassalho et al., 2017).

[Insert Table 6]

In general, the linear model showed advantages in the adjustment up to the recurrence period of 20 years, while the recurrence periods of 50 and 100 years the power model had the best performance. Another important issue is that Da and P_{eq} had the best performance as predictor variables.

Table 8 and Table 9 show the regionalization models for the annual and monthly average streamflow, respectively. Just like the maximum streamflow, the linear and non-linear models performed similarly for the average annual streamflow regionalization, however, the power model had the best performance. The statistical indexes were considered excellent in terms of precision ($R^2 > 0.98$) and accuracy ($d > 0.99$), with relatively low errors (RMSE < 8.0 $m^3\ s^{-1}$) (Pruski et al., 2013, 2016; Cecilio et al., 2018). For monthly average streamflow, nonlinear (power) models should be preferred due to its better statistical performance in the streamflow representation.

The nonlinear regression models for predicting the average monthly streamflow had significant coefficients by the t-test, and the "a" coefficient followed the streamflow seasonality, decreasing its value between March and April until reaching its minimum value in August. In September and October, its value increases again, reaching the maximum value in December, January, or February. The "b" coefficient also showed such a pattern indicating a potent relation of streamflow reduction during the dry period, with Da, $P_{eq,}$ or P_{eq750}. The predictor variables with the best performance for the average streamflow, as well as for the maximum streamflow, were Da and P_{eq}.

[Insert Table 8]

[Insert Table 9]

Table 10 shows the linear and non-linear models for $Q_{7,10}$ regionalization and the streamflow permanence curve (Q_{95}, Q_{90}, $Q_{80,}$ and Q_{50}) in the Sapucaí River watershed. In general, the linear and non-linear models were able to predict minimum streamflow with excellent precision ($R^2 > 0.98$), accuracy ($d > 0.99$), and errors that were considered acceptable (RMSE <1.4 $m^3\ s^{-1}$). All parameters from the linear regression but the parameter "a" were significant. The parameter "a" from the linear regression indicates the position where the line intersects the Y-axis and, when it is statistically equal to 0, the line passes through the origin, indicating linearity between the predictor variables and the minimum reference streamflow.

[Insert Table 10]

4. DISCUSSION

Basically, the highest streamflow was found in gauges with large drainage areas (Da), while in rivers with small Da the streamflow tended to be lower. This confirms the relationship between streamflow and Da, which justifies the use of this variable as a streamflow predictor in regionalization studies. Another important indicator is the relationship between streamflow and frequently flooded areas, such as the gauges 61350000 and 6130500 (close to the region of Santa Rita do Sapucaí and São Sebastião da Bela Vista), and the gauges 6127200 and 6128500 (located in Itajubá). The streamflow in these gauges was considered of intermediate magnitudes. It indicates that other factors affect the flood regime in these regions, such as the physiography and the difference in slope (Almeida et al. 2017). Also, the trend analysis, carried out for the gauges (61250000 and 61285000) near the city of Itajubá, did not show increasing trends for maximum and average streamflow (Almeida, Silva, Cecílio, Abreu, & Fraga, 2019), which corroborates with the physiographic conditions as the main flood propensity factor.

The parameters of the PDFs were correlated with the predictive variables and the reference streamflows and it is interesting for the physical meaning of probability distributions. One major problem faced by water engineers is the determination of the most suitable form of an extreme value (maximum or minimum) probability distribution of the flood, and the approximation of parameters of the distribution. However, the parameter values that give the maximum likelihood function among so many other possible sample series of the population are considered the most suitable ones for that sample series (Langat, Kumar, & Koech, 2019).

The GEV distribution is characterized by a good representation of positive asymptotic series, as is the maximum streamflow (Cassalho et al., 2017; Castellarin, 2007; Guse, Hofherr, & Merz, 2010). For the hydrological series of maximum annual daily rainfall, which also shows a positive asymptotic pattern, the PDF GEV was also efficient in representing this data (Abreu et al., 2018; Beskow, Caldeira, Mello, Faria, & Guedes, 2015). Therefore, it is a promising PDF in hydrological studies of this nature. Also, the GEV has three parameters that can be adjusted, while the other PDFs only have two parameters, hence the GEV adjustment is more flexible adjustment (Lyra et al., 2006). The normal distribution's worst performance can be justified by its symmetry characteristic and the bell shape. The normal distribution is considered to have extremely limited flexibility in terms of asymmetry which is ideal to represent random variables that fluctuate symmetrically around the mean.

The normal distribution was tested only in a few streamflow studies, due to its simplicity (it has two parameters) and, mainly, due to its characteristic of symmetry of the values around the mean. This characteristic is not expected for most hydrological variables, including streamflow. Finkler et al. (2015), tested the normal distribution to obtain $Q_{7,10}$ and it showed only 12.5% of inadequacies in representing the data by the χ^2 goodness-of-fit test and 100% adherence in the KS and AD tests, in the Arroio Belo River watershed in the of state of Rio Grande do Sul. However, some studies have shown an inadequate or inferior performance of the normal distribution compared to other PDFs (Langat et al., 2019; Modarres, 2008), which highlights the need to verify the best PDF for each place. Another distributions with suitable fit in low flows are: Gamma (Konrad & Rumsey, 2019), Weibull, Gumbel and GEV (Langat et al., 2019).

Comparatively, the power regression showed a slightly better adjustment for the minimum reference streamflow prediction ($Q_{7,10}$), and it is the most used in regionalization studies (Pruski et al., 2016; Cecilio et al., 2018).

Another important observation is that $Q_{7,10}$ always had lower values than Q_{95} the minimum reference streamflow, therefore the $Q_{7,10}$ is stricter in terms of defining the limit for water withdrawals. Other studies have also found the same results (Ouyang, 2012; Serrano et al., 2020).

The good performance of Da as a streamflow predictor variable (maximum, average and minimum) is relevant because it is one of the most used variables in streamflow regionalization (Razavi & Coulibaly, 2013) due to its close relationship with the streamflow, as shown in Figure 6. However, some researches indicate that regionalization equations exclusively conditioned to Da may not reflect the effect of the variation in rainfall along the watershed (Pruski et al., 2013, 2015, 2016; Cecilio et al., 2018). As the rainfall variability was small (between 1380 and 1522 mm; the standard deviation of 63 mm) in the watershed under study, the better performance of Da as a predictor variable is justified. The streamflow gauges used were in drainage areas of different sizes, which may have contributed to the Da's good performance a predictor variable.

The consequences of using Da as a predictor variable in streamflow regionalization studies would be the overestimation of the streamflow in the upper areas of the watershed, where the drainage areas are smaller than the Da used to establish regionalization equations (Silva Junior, Bueno, Tucci, & Castro, 2003). To minimize such effect, Pruski et al. (2015) indicate the use of a threshold value, which is the maximum specific streamflow estimated in the streamflow gauges used in the regionalization study.

For maximum streamflow, the better performance of the equations using Da is explained by the fact that the average annual rainfall has little relation to the maximum annual streamflow. The Q_{max} is the watershed's hydrological response to a small set of specific rainfall events, usually the maximum rainfall of one day, or the accumulated rainfall of five days (Avila, Justino, Wilson, Bromwich, & Amorim, 2016), associated with the hydraulic characteristics of the watercourses and land use. For average and minimum streamflow, the effect of average annual rainfall is more relevant, although in this study only P_{eq} showed a similar statistical performance as the Da.

Therefore, the traditional method is efficient for the regionalization of maximum, average, and minimum streamflow in the Sapucaí river watershed. Despite the similar performance of the first-order linear and non-linear power regression models, the second can be considered having a better performance. Also, with the easily obtained data such as Da and P_{eq}, reference streamflow from the Sapucaí River watershed can be reliably estimated. However, it is recommended, for the Sapucaí River watershed, the use of Da because it had better statistical performance and the low rainfall representativeness in the watershed due to the scarce rainfall data.

5. CONCLUSIONS

The goodness-of-fit tests used for the frequency analysis of the reference streamflow were different in terms of rigor in accepting the null hypothesis of adherence of the probability density function (PDF) to the data set. The Kolmogorov-Smirnov and χ^2 tests were the most permissive in accepting the null hypothesis and the most rigorous ones were the Filliben and Shapiro-Wilk tests. The probability density function with the best performance in representing maximum annual streamflow was generalized of extreme values (GEV), while for the representation of minimum streamflow was the normal distribution.

The streamflow regionalization models, linear and non-linear, with the drainage area (Da) or with the streamflow equivalent to the rainfall volume (considering or not the abstractions) as predictive variables, were efficient in estimating the reference streamflow. Despite the similar performance, the nonlinear power regressions were superior to the linear regression model, as well as the drainage area, and the streamflow equivalent to the rainfall volume (without abstractions) should be preferred in regionalization due to modest statistical superiority.

.

DATA AVAILABILITY

The data that support the findings of this study are available in National Water Agency (ANA) - HidroWeb – Hydrological Information System platform at http://www.snirh.gov.br/hidroweb/publico/medicoes_historicas_abas.jsf. and National Institute of Meteorology (INMET) at https://portal.inmet.gov.br/?r=estacoes/estacoesAutomaticas.

Some or all data, models, or code generated or used during the study are available in a repository online in accordance with funder data retention policies.

The daily streamflow records were obtained from the hydrometeorological database of the National Water Agency (ANA), through the HidroWeb – Hydrological Information System platform:

http://www.snirh.gov.br/hidroweb/publico/medicoes_historicas_abas.jsf.

The daily rainfall records were obtained from the hydrometeorological database of the National Institute of Meteorology (INMET), through the platform:

https://portal.inmet.gov.br/?r=estacoes/estacoesAutomaticas

Some or all data, models, or code generated or used during the study are available from the corresponding author by request.

The models and R codes, for example, can be made available.

REFERENCES

Abreu, M. C., Cecílio, R. A., Pruski, F. F., dos Santos, G. R., Almeida, L. T., & Zanetti, S. S. (2018). Critérios para escolha de distribuições de probabilidades em estudos de eventos extremos de precipitação. *Revista Brasileira de Meteorologia,* 33(4), 601–613. https://doi.org/10.1590/0102-7786334004

Almeida, L. T., Silva, F. B., Cecílio, R. A., Abreu, M. C., & Fraga, M. D. S. (2019). Análise do comportamento da vazão e precipitação na influência de enchentes na bacia hidrográfica a montante da cidade de Itajubá. *Revista Augustus,* 24, 124–145.

Almeida, L. T., Abreu, M. C., Fraga, M. D. S., Silva, D. D., & Cecílio, R. A. (2017). Aspectos morfométricos relacionados ao estudo de enchentes na Bacia do Rio Sapucaí Minas Gerais. *Nativa,* 5, 169–174. http://dx.doi.org/10.5935/2318-7670.v05n03a03

Alvares, C. A., Stape, J. L., Sentelhas, P. C., Gonçalves, J. L. M., & Sparovek, G. (2013). Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift,* 22, 711–728. https://doi.org/10.1127/0941-2948/2013/0507

Amorim, J. D. S., Junqueira, R., Mantovani, V. A., Viola, M. R., Mello, C. R., & Bento, N. L. (2020). Streamflow regionalization for the Mortes River Basin upstream from the Funil Hydropower Plant MG. *Ambiente Agua*, 15(3), e2495. https://doi.org/10.4136/ambi-agua.2495

Andrade, M. A., Mello, C. R., & Beskow, S. (2013). Simulação hidrológica em uma bacia hidrográfica representativa dos Latossolos na região Alto Rio Grande MG. *Revista Brasileira de Engenharia Agrícola e Ambiental,* 17(1), 69–76. https://doi.org/10.1590/S1415-43662013000100010

Avila, A., Justino, F. B., Wilson, A., Bromwich, D., & Amorim, M. (2016). Recent precipitation trends flash floods and landslides in southern Brazil. *Environmental Research Letters,* 11, 114029. https://doi.org/10.1088/1748-9326/11/11/114029

Baena, L. G. N., Silva, D. D. D. A., Pruski, F. F., & Calijuri, M. L. (2004). Regionalização de vazões com base em modelo digital de elevação para a bacia do rio Paraíba do Sul. *Engenharia Agrícola,* 24(3), 612–624. http://dx.doi.org/10.1590/S0100-69162004000300013

Barros, C. G. D., Pessoa, F. C. L., Santana, L. R., Lopes, Y. K. L., & Costa, C. E. A. S. (2018). Vazão mínima $Q_{7,10}$ no Amapá estimada por modelos probabilísticos. *Engenharia Agrícola,* 26(3), 284–294. https://doi.org/10.13083/reveng.v26i3.930

Beskow, S., Caldeira, T. L., Mello, C. R., Faria, L. C., & Guedes, H. A. S. (2015). Multiparameter probability distributions for heavy rainfall modeling in extreme southern Brazil. *Journal of Hydrology: Regional Studies*, 4, 123–133. https://doi.org/10.1016/j.ejrh.2015.06.007

Beskow, S., Mello, C. R., Faria, L. C., Simões, M. C., Caldeira, T. L., & Nunes, G. S. (2014). Índices de sazonalidade para regionalização hidrológica de vazões de estiagem no Rio Grande do Sul. *Brasileira de Engenharia Agrícola e Ambiental*, 18(7), 748–754. http://dx.doi.org/10.1590/S1415-43662014000700012

Beskow, S., Norton, L. D., & Mello, C. R. (2013). Hydrological prediction in a tropical watershed dominated by Oxisols using a distributed hydrological model. *Water Resources Management,* 27, 341–363. https://doi.org/10.1007/s11269-012-0189-8

Cassalho, F., Beskow, S., Vargas, M. M., Moura, M. M., Ávila, L. F., & Mello, C. R. (2017). Hydrological regionalization of maximum stream flows using an approach based on L-moments. *Revista Brasileira de Recursos Hídricos*. 22, e27. https://doi.org/10.1590/2318-0331.021720160064

Castellarin, A. (2007). Probabilistic envelope curves for design flood estimation at ungauged sites. *Water Resourses Research*, 43, 1–12. https://doi.org/10.1029/2005WR004384

Cecílio, R. A., Zanetti, S. S., Gasparini, K. A., & Catrinck, C. N. (2018). Avaliação de métodos para regionalização das vazões mínimas e médias na bacia do rio Itapemirim. *Scientia Agraria*, 19, 122–132. http://dx.doi.org/10.5380/rsa.v19i2.52726

Costa, M. D. S., Alfenas, U. F., Beijo, L. A., Alfenas, U. F., Avelar, F. G., & Alfenas, U. F. (2019). Comparação de distribuições de probabilidades na previsão de vazões máximas do reservatório de Furnas. *Revista Brasileira de Agricultura Irrigada*, 13(1), 3190–3202. DOI: 10.7127/rbai.v13n100893

Duraes, M. F., & Mello, C. R. (2016). Distribuição espacial da erosão potencial e atual do solo na Bacia Hidrográfica do Rio Sapucaí MG. *Engenharia Sanitária e Ambiental*. 21(4), 677–685. https://doi.org/10.1590/s1413-41522016121182

Duraes, M. F., Mello, C. R., & Beskow, S. (2015). Estresse hidrológico: aplicação às bacias dos rios Paraopeba e Sapucaí Minas Gerais. *Revista Brasileira de Recursos Hídricos*. 20(2), 352–359. https://doi.org/10.21168/rbrh.v20n2.p352-359

Elesbon, A. A. A., Silva, D. D., Sediyama, G. C., Guedes, H. A. S., Ribeiro, C. A. A. S., Ribeiro, C. B. M. (2015). Multivariate statistical analysis to support the minimum streamflow regionalization. *Engenharia Agrícola*. 35(5), 838–851. https://doi.org/10.1590/1809-4430-Eng.Agric.v35n5p838-851/2015

ELETROBRAS - (1985). Metodologia para regionalização de vazões in: Centrais Hidrelétricas Brasileiras. Eletrobras Rio de Janeiro.

Finkler, N. R., Mendes, L. A., Schneider, H. E. M., Bortolin, T. A., & Schneider, V. E. (2015). Comparação de funções de distribuição de probabilidades na determinação de vazão mínima anual e sazonal. *Scientia cum Industria*. 3(2), 42–49. http://dx.doi.org/10.18226/23185279.v3iss2p42

Granemann, A. R. B., Mine, M. R. M., & Kaviski, E. (2018). Frequency analysis of minimum flows. *Revista Brasileira de Recursos Hídricos*. 23(e17), 1–14. https://doi.org/10.1590/2318-0331.0318170080

Guse, B., Hofherr, T. H., & Merz, B. (2010). Introducing empirical and probabilistic regional envelope curves into a mixed bounded distribution function. *Hydrology and Earth System Sciences*. 14, 2465–2478. https://doi.org/10.5194/hess-14-2465-2010

Hannaford, J., Buys, G., Stahl, K., & Tallaksen, L. M. (2013). The influence of decadal-scale Syst, variability on trends in long European streamflow records. *Hydrology and Earth System Sciences*.17: 2717–2733. https://doi.org/10.5194/hess-17-2717-2013

IBGE – Instituto Brasileiro de Geografia e Estatística. (2010). IBGE Cidades. <https://cidades.ibge.gov.br/>Accessed September 01 2020

IBGE – Instituto Brasileiro de Geografia e Estatística. (2017). IBGE Cidades. <https://cidades.ibge.gov.br/>Accessed September 01 2020

IBGE – Instituto Brasileiro de Geografia e Estatística. (2019). Estimativas da população residente para os municípios e para as unidades da federação com data de referência em 1º de julho de 2019: notas metodológicas. Rio de Janeiro.

Konrad, C., & Rumsey, C. (2019). Estimating minimum streamflow from measurements at ungauged sites in regions with streamflow−gauging networks. *Hydrological Process*. 33, 2057-2067. https://doi.org/10.1002/hyp.13452

Koutroulis, A. G., Papadimitriou, L. V., Grillakis, M. G., Tsanis, I. K., Warren, R., & Betts, R. A. (2019). Global water availability under high-end climate change: A vulnerability-based assessment. *Global and Planetary Change*. 175, 52-63. https://doi.org/10.1016/j.gloplacha.2019.01.013

Langat, P. K., Kumar, L., & Koech, R. (2019). Identification of the most suitable probability distribution models for maximum minimum and mean streamflow. *Water*. 11, 1–24. https://doi.org/10.3390/w11040734

Lisboa, L., David, D., Moreira, M. C., Silva, A. D. J., & Uliana, E. M. (2019). Sistema para análise das outorgas de captação de água e diluição de efluentes na bacia do rio Piracicaba (MG). *Engenharia Sanitária e Ambiental*. 24, 929–937. https://doi.org/10.1590/S1413-41522019183919

Lopes, T. R., Prado, G., Zolin, C. A., Paulino, J., & Antoniel, L. S. (2016). Regionalização de vazões máximas e mínimas para a bacia do rio Ivaí. *Irriga* 21, 188–201. https://doi.org/10.15809/irriga.2016v21n1p188-201

Lyra, G. B., Lozada, Garcia, B. I., Piedade, S. M. S., Sediyama, G. C., & Sentelhas, P. C. (2006). Regiões homogêneas e funções de distribuição de probabilidade da precipitação pluvial no Estado de Táchira Venezuela. *Pesquisa Agropecuária Brasileira*. 41(2), 205–215. https://doi.org/10.1590/S0100-204X2006000200004

Maciel, A. L., Vieira, E. M., Monte Mor, R. C., & Vasques, A. C. (2019). Regionalização e espacialização de vazões de permanência: estudo aplicado na bacia Rio Piracicaba-MG. Revista Brasileira de Climatologia. 15(24), 114–133. http://dx.doi.org/10.5380/abclima.v24i0.58420

Martins, F. B., Gonzaga, G., Santos, D. F., & Reboita, M. S. (2018). Classificação climática de Köppen e de Thornthwaite para Minas Gerais: cenário atual e projeções futuras. *Revista Brasileira de Climatologia*. 11, 129–156. https://doi.org/10.5380/abclima.v1i0.60896

Matos, A. J. S., Pioltine, A., Mauad, F. F., & Barbosa, A. A. (2011). Metodologia para a caracterização do coeficiente de Manning variando na seção transversal e ao longo do canal: estudo de caso bacia do Alto Sapucaí. *Revista Brasileira de Recursos Hídricos*. 16, 21–28. https://doi.org/10.21168/rbrh.v16n4.p21-28

Matos, T. S., Uliana, E. M., Martins, C. A. S., & Rapalo, L. M. C. (2020). Regionalization of maximum minimum and mean stream flows for the Juruena River basin, Brazil. *Ambiente e Água*. 15(3), 1-18. https://doi.org/10.4136/ambi-agua.2418

Mediero, L., Kjeldsen, T. R., Macdonald, N., Kohnova, S., Merz, B., Vorogushyn, S., Wilson, D., Perdigão, R. A. P., Roald, L. A., Salinas, J. L., Toumazis, A. D., Lang, M., Madsen, H., & Onus, G. (2015). Identification of coherent flood regions across Europe by using the longest streamflow records. *Journal of Hydrology*. 528, 341–360. https://doi.org/10.1016/j.jhydrol.2015.06.016

Millard, S. P. (2013). EnvStats: An R Package for Environmental Statistics.

Modarres, R. (2008). Regional frequency distribution type of low flow in North of Iran by L-moments. *Water Resources Management*. 22, 823–841. https://doi.org/10.1007/s11269-007-9194-8

Naghettini, M., & Pinto, E. J. A. (2007). Hidrologia Estatística. CPRM ed Belo Horizonte. Available in: <http://www.cprm.gov.br/publique/Hidrologia/Mapas-e-Publicacoes/Livro-%22Hidrologia-Estatistica%22-981.html>

Ouyang, Y. (2012). A potential approach for low flow selection in water resource supply and management. *Journal of Hydrology*. 454–455(6), 56–63. https://doi.org/https://doi.org/10.1016/j.jhydrol.2012.05.062

Piol, M. V. A., Reis, J. A. T., Caiado, M. A. C., & Mendonça, A. S. F. (2019). Performance evaluation of flow duration curves regionalization methods. *Revista Brasileira de Recursos Hídricos*. 24(28), 1–13. https://doi.org/10.1590/2318-0331.241920170202

Pruski, F. F., Rodriguez, R. D. G., Pruski, P. L., Nunes, A. D. E. A., & Rego, F. S. (2016). Extrapolation of regionalization equations for long-term average flow. *Engenharia Agrícola*. 36(5), 830–838. https://doi.org/10.1590/1809-4430-Eng.Agric.v36n5p830-838/2016

Pruski, F. F., Nunes, A. A., Pruski, P. L., & Rodriguez, R. G. (2013). Improved regionalization of streamflow by use of the streamflow equivalent of precipitation as an explanatory variable. *Journal of Hydrology*. 476(7), 52–71. https://doi.org/10.1016/j.jhydrol.2012.10.005

Pruski, F. F., Rodriguez, R. D. G., & Nunes, A. A. (2015). Low-flow estimates in regions of extrapolation of the regionalization equations: a new concept. *Engenharia Agrícola*. 35(5), 808–816. https://doi.org/10.1590/1809-4430-Eng.Agric.v35n5p808-816/2015

R Team Core (2018). R: A language and environment for statistical computing.

Razavi, T., & Coulibaly, P. (2013). Streamflow Prediction in Ungauged Basins: Review of Regionalization Methods. *Journal of Hydrologic Engineering*. 18, 958–975. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000690

Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). mclust 5: Clustering classification and density estimation using Gaussian finite mixture models.

Serrano, L. O., Ribeiro, R. B., Borges, A. C., & Pruski, F. F. (2020). Low-Flow Seasonality and Effects on Water Availability throughout the River Network. *Water Resources Management*. 1–16. https://doi.org/10.1007/s11269-020-02499-3

Silva, A. M., Oliveira, P. M., Mello, C. R., & Pierangeli, C. (2006). Vazões mínimas e de referência para outorga na região do Alto Rio Grande Minas Gerais. *Revista Brasileira de Engenharia Agrícola e Ambiental*. 10(2), 374–380. https://doi.org/10.1590/S1415-43662006000200019

Silva Junior, B. B., Bueno, E. O., Tucci, C. E. M., & Castro, N. M. R. (2003). Extrapolação espacial na regionalização da vazão. *Revista Brasileira de Recursos Hídricos*. 8, 21–37. DOI: 10.21168/rbrh.v8n1.p21-37

Sousa, H. T., Pruski, F. F., Bof, L. H. N., Cecon, P. R., & Sousa, J. R. C. (2009). Sistema Computacional para Análises Hidrológicas SISCAH 1.0 - Grupo de Pesquisas em Recursos Hídricos Universidade Federal de Viçosa.

Wolff, W., Duarte, S., & Mingoti, R. (2014). Nova metodologia de regionalização de vazões estudo de caso para o Estado de São Paulo. *Revista Brasileira de Recursos Hídricos*. 19, 21–33. https://doi.org/10.21168/rbrh.v19n4.p21-33.

APPENDICES

Attachment 1 Parameters of location, scale and threshold for the Probability density function for the maximum annual streamflow.

fdp/ parameters	Gamma		Weibull		Normal		Log-normal		Gumbell		GEV		
	Shape	Scale	Shape	Scale	Shape	Scale	Shape	Scale	Shape	Scale	Location	Shape	Scale
Station	Q_{max} ($m^3 s^{-1}$)												
61343000	7.787	4.243	3.252	36.714	33.041	10.757	3.432	0.398	27.774	10.668	28.942	10.488	0.209
61295000	2.948	10.581	1.851	35.247	31.190	17.817	3.261	0.630	23.011	13.795	22.596	13.469	-0.056
61565000	3.279	16.543	1.681	61.200	54.242	35.891	3.833	0.560	40.802	21.642	39.079	20.307	-0.141
61460000	5.199	35.771	2.416	209.492	185.972	80.156	5.126	0.473	149.436	66.616	151.723	67.340	0.062
61350000	3.367	46.321	1.832	176.622	155.981	91.678	4.894	0.565	117.145	62.305	112.518	58.679	-0.143
61250000	4.250	4.868	2.142	23.494	20.689	10.395	2.907	0.495	15.965	7.650	15.118	6.843	-0.220
61272000	7.391	15.857	3.042	130.939	117.195	40.907	4.695	0.397	97.618	37.419	100.797	37.796	0.160
61370000	7.027	10.868	3.641	84.297	76.372	23.471	4.263	0.446	64.140	26.242	70.257	25.319	0.466
61285000	13.187	5.472	4.579	79.183	72.160	18.567	4.240	0.289	62.609	18.664	67.072	19.748	0.445
61305000	10.125	22.020	3.985	246.064	222.957	63.853	5.357	0.337	190.145	65.675	202.339	66.474	0.347
61510000	4.771	84.132	2.139	454.163	401.359	197.358	5.886	0.475	318.465	140.832	315.332	139.454	-0.041

61537000	6.053	92.335	2.732	627.686	558.891	216.383	6.241	0.440	456.408	191.551	469.790	194.679	0.130
61390000	8.401	6.442	3.865	59.755	54.122	16.073	3.931	0.388	45.770	17.468	50.035	17.349	0.480
61320000	4.861	13.351	2.555	73.197	64.898	27.450	4.066	0.492	51.566	24.331	54.288	25.693	0.205

Attachment 2 Parameters of location, scale and threshold for the Probability density function for the minimum annual streamflow.

fdp/parameters	Gamma		Weibull		Normal		Log-normal		Gumbell		GEV		
	Shape	Scale	Shape	Scale	Shape	Scale	Shape	Scale	Shape	Scale	Location	Shape	Scale
Station	Q_7 (m^3 s^{-1})												
61343000	18.130	0.121	4.738	2.400	2.196	0.511	0.759	0.239	1.947	0.455	2.003	0.484	0.228
61295000	9.991	0.089	3.194	0.993	0.889	0.292	-0.168	0.317	0.756	0.224	0.749	0.219	-0.057
61565000	9.918	0.289	3.651	3.185	2.866	0.879	1.002	0.329	2.431	0.792	2.550	0.860	0.271
61460000	16.421	0.878	4.679	15.768	14.421	3.438	2.638	0.255	12.712	3.322	13.276	3.458	0.322
61350000	14.893	0.607	4.411	9.930	9.040	2.296	2.168	0.265	7.907	2.065	8.251	2.266	0.302
61250000	31.641	0.054	5.648	1.831	1.705	0.303	0.517	0.187	1.555	0.319	1.590	0.306	0.208
61272000	26.273	0.329	6.980	9.222	8.633	1.513	2.136	0.209	7.809	1.897	not fit	not fit	not fit
61370000	13.408	0.319	3.986	4.720	4.282	1.156	1.417	0.279	3.723	1.031	3.824	1.071	0.180
61285000	10.455	0.479	3.918	5.520	5.007	1.412	1.562	0.335	4.292	1.465	4.519	1.444	0.296
61305000	11.764	1.755	4.279	22.676	20.652	5.471	2.985	0.315	17.844	5.807	18.876	5.699	0.343
61510000	16.521	1.973	4.438	35.694	32.597	7.904	3.454	0.252	28.750	7.265	29.565	7.495	0.209
61537000	27.846	1.796	5.837	53.970	50.013	9.478	3.894	0.193	45.352	8.440	46.623	9.207	0.272
61390000	11.453	0.269	3.688	3.403	3.078	0.887	1.080	0.306	2.646	0.814	2.724	0.835	0.172
61320000	13.676	0.325	4.680	4.864	4.445	1.109	1.455	0.287	3.870	1.160	4.137	1.183	0.436

TABLES

TABLE 1 Stream gauge stations and rain gauge stations selected for the study

Stream gauge stations					
Code	Station name	River	Latitude	Longitude	Da (km^2)
61343000	Bairro do Analdino	Capivari	-22.550	-45.883	247
61295000	Brasópolis	Vargem Grande	-22.467	-45.622	156
61565000	Cachoeira Poço Fundo	Machado	-21.783	-45.124	349
61460000	Conceição do Rio Verde	Verde	-21.883	-45.079	1840
61350000	Conceição dos Ouros	Sapucaí-Mirim	-22.400	-45.791	1310
61250000	Fazenda da Guarda	Sapucaí	-22.683	-45.480	109
61272000	Itajubá	Sapucaí	-22.433	-45.450	870
61370000	Ponte do Rodrigues	Itaim	-22.350	-45.854	676
61285000	São João de Itajubá	Lourenço Velho	-22.367	-45.448	560
61305000	Santa Rita do Sapucaí	Sapucaí	-22.250	-45.709	2810
61510000	Três Corações	Verde	-21.700	-45.248	4180
61537000	UHE Furnas rio Verde	Verde	-21.600	-45.489	6300
61390000	Vargem do Cervo	Cervo	-22.117	-45.918	468
61320000	São Bento do Sapucaí	Sapucaí-Mirim	-22.683	-45.735	475
Rain gauge stations					
Code (OMM)	Station name	Altitude (m)	Latitude	Longitude	
83714	Campos do Jordão	1642	-22.75	-45.6	-
83032	Lambari	878.45	-21.94	-45.31	-
83687	Lavras	918.84	-21.75	-45.00	-
83683	Machado	873.35	-21.68	-45.94	-
83736	São Lourenço	953.20	-22.10	-45.01	-

Da = Drenage area

TABLE 2 The probability density functions and cumulative distribution function for the probability density frequency (PDFs)

PDF	Probability Density Function	Cumulative Distribution Function	Parameters	Observations
Gamma	$f_x(x)=\frac{\left(x/\theta\right)^{\eta-1}\exp\left(-x/\theta\right)}{\theta\Gamma(\eta)}$	$F_x(x)=\int_0^\infty\frac{\left(x/\theta\right)^{\eta-1}\exp\left(-x/\theta\right)}{\theta\Gamma(\eta)}\,dx$	Θ = scale η = shape	For: X, Θ and η > 0 $\Gamma(\eta)=\int_0^\infty x^{\eta-1}e^{-x}dx$
Weibull	$f_x(x)=\frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha\text{-}1}\exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right]$	$Fx(x)=1-\exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right]$	α = scale β = shape	
Normal	$f_x(x)=\frac{1}{\sqrt{2\pi\theta_2^2}}\exp\left\{-\frac{1}{2}\left[\frac{(x-\theta_1)}{\theta_2}\right]^2\right\}$	$F_x(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi\theta_2^2}}\exp\left\{-\frac{1}{2}\left[\frac{(x-\theta_1)}{\theta_2}\right]^2\right\}dx$	$E[X]=\mu=\Theta_1$ $Var[X]=\sigma^2=\Theta_2$	For: -∞ < X < ∞ μ = mean σ = standart derivation
Log-normal	$f_x(x)=\frac{1}{x\sigma_{\ln(x)}\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left[\frac{\ln(x)-\mu_{\ln(x)}}{\sigma_{\ln(x)}}\right]^2\right\}$	$F_x(x)=\int_{-\infty}^{x}\frac{1}{x\sigma_{\ln(x)}\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left[\frac{\ln(x)-\mu_{\ln(x)}}{\sigma_{\ln(x)}}\right]^2\right\}dx$	$E[X]=\mu=\Theta_1$ $Var[X]=\sigma^2=\Theta_2$	For: -∞ < X < ∞ μ = mean σ = standart derivation
Gumbel	$f_x(x)=\frac{1}{\alpha}\exp\left[-\frac{x-\beta}{\alpha}-\exp\left(-\frac{x-\beta}{\alpha}\right)\right]$	$F_x(x)=\exp\left[-\exp\left(-\frac{x-\beta}{\alpha}\right)\right]$	α = scale β = shape	For: -∞ < X < ∞ -∞ < β < ∞ α < ∞

GEV	$f_x(x)=\frac{1}{\alpha}\left[1-k\left(\frac{x-\beta}{\alpha}\right)^{1/(k-1)}\exp\left\{-\left[1-k\left(\frac{x-\beta}{\alpha}\right)\right]^{1/k}\right\}\right]$	$F_x(x)=\exp\left\{-\left[1-k\left(\frac{x-\beta}{\alpha}\right)\right]^{1/k}\right\}$	α = scale β = shape k = position

TABLE 3 The goodness-of-fit test and the statistic of the test.

The goodness-of-fit test	Statistics of the test
KS	$\lvert\Delta F\rvert_{max}$
χ^2	$\chi^2 = \sum_{i=1}^{n} \frac{\left(f_{\text{obs-i}} - f_{\text{theoretical-i}}\right)^2}{f_{\text{theoretical-i}}}$
AD	$AD^2 = -N - \sum_{i=1}^{N} \frac{(2i-1)\left[\ln(P1) + \ln(P2)\right]}{N}$
Fi	$r_{Fi} = \frac{\sum_{i=1}^{N}\left(X_i - \overline{X}\right)\cdot\left(W_i - \overline{W}\right)}{\sqrt{\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2\cdot\sum_{i=1}^{N}\left(W_i - \overline{W}\right)}}$
CVM	$CVM^2 = \frac{1}{2N} + \sum_{i=1}^{N}\left[p(i) - \frac{2i-1}{2N}\right]^2$
SW	$W = \frac{\sum_{i=1}^{N}\left(p_{(i)} X_i\right)^2}{\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2}$

$\lvert\Delta F\rvert_{max}$ is the value of the KS test statistic, obtained through the largest difference between the cumulative, empirical and theoretical functions; $f_{obs\text{-}i}$ is the frequency observed in the i-th class; $f_{theoretical\text{-}i}$ is the theoretical frequency in the i-th class; ie; P1 is the probability of non-exceedance calculated by the probability distribution with the data in ascending order, P2 is the probability of exceedance calculated by the probability distribution with the data in descending order and N is the sample size; X_i is the quantile observed in the i-th observation, $\overline{X}$ is the average of the observed quantiles, W_i is the theoretical quantile in the i-th observation, $\overline{W}$ is the average of the estimated quantiles; $p(i) = \Phi([X(i) - \overline{X}]/s)$; Φ is the cumulative distribution of the density function of X(i) is the observation at the i-th position and s is the mean standard deviation of the observed values.

TABLE 4 Monthly and annual long-term average streamflow ($m^3\ s^{-1}$) of each stream gauge station in the Sapucaí River basin

Estações	Q_{jan}	Q_{feb}	Q_{mar}	Q_{apr}	Q_{may}	Q_{jun}	Q_{jul}	Q_{aug}	Q_{sep}	Q_{ouc}	Q_{nov}	Q_{dec}	Q_{avg}
61343000	11.1	11.1	9.2	6.9	5.2	4.3	3.7	3.0	3.0	3.5	4.1	5.8	5.9
61295000	6.5	5.8	4.3	2.9	2.2	1.8	1.6	1.2	1.3	1.6	2.0	3.2	2.9
61565000	16.4	14.8	12.7	9.1	6.7	5.7	4.8	4.0	4.2	5.0	6.5	10.5	8.4
61460000	78.9	68.7	55.1	41.4	30.7	25.8	22.1	18.6	19.0	21.3	28.0	45.3	37.9
61350000	44.6	44.3	36.0	27.4	20.6	17.4	15.1	12.2	12.3	14.9	17.6	25.7	24.0
61250000	5.3	5.6	4.6	3.8	3.2	2.8	2.6	2.2	2.2	2.5	2.9	3.9	3.5
61272000	34.0	32.4	28.2	22.2	18.0	15.4	13.0	10.8	11.3	13.5	16.3	21.8	19.8
61370000	27.0	27.4	21.2	15.2	11.1	9.5	8.1	6.3	6.8	8.3	9.6	13.7	13.7
61285000	21.7	21.0	17.9	13.8	10.6	9.1	7.7	6.5	6.6	7.4	9.2	13.7	12.1
61305000	96.9	92.5	75.9	57.7	45.7	39.7	33.3	27.6	28.8	32.7	41.0	58.8	52.5
61510000	176.6	153.1	126.0	94.1	70.0	59.4	50.5	42.6	43.3	48.7	62.8	99.8	85.6
61537000	271.5	267.0	207.8	152.7	113.5	94.7	79.9	65.6	67.5	81.7	97.4	143.2	136.9
61390000	20.3	19.4	15.9	10.9	7.9	6.5	5.3	4.3	4.7	6.1	7.7	11.5	10.0
61320000	17.5	16.3	13.6	11.2	8.5	7.4	6.7	5.6	5.7	7.0	7.9	11.3	9.9
Mean	59.2	55.7	44.9	33.5	25.3	21.4	18.2	15.0	15.5	18.1	22.4	33.5	30.2
Sd	77.0	73.3	57.7	42.6	31.7	26.6	22.4	18.6	19.1	22.6	27.5	41.4	38.3
Skewness	2.1	2.2	2.1	2.1	2.1	2.0	2.0	2.0	2.0	2.1	2.0	1.9	2.1
Kurtosis	4.0	5.1	4.6	4.3	4.1	3.9	3.8	3.6	3.7	4.4	3.7	3.2	4.2

Sd = Standard derivation

TABLE 5 Kurtosis and Skewness for maximum streamflow, and minimum streamflow of seven consecutive days

61305000	0.164	-0.292	0.526	-0.289
61510000	4.258	1.757	0.054	0.306
61537000	1.170	0.652	-0.854	0.149
61390000	0.346	-0.473	1.000	0.410
61320000	-0.487	0.292	0.025	-0.404

TABLE 6 PDF percentage of adherence in each goodness-of-fit test

PDF	KS	X^2	AD	CVM	Fi	SW
Maximum streamflow						
Gamma	100	93	93	86	71	71
Weibull	100	93	71	79	79	79
Normal	100	79	64	64	57	57
Log-normal	100	93	71	71	50	57
Gumbel	100	93	79	79	57	64
GEV	100	93	100	100	100	100
Minimum streamflow of seven consecutive days						
Gamma	100	100	86	93	79	86
Weibull	100	93	93	93	93	93
Normal	100	100	100	100	86	100
Log-normal	100	100	71	71	64	64
Gumbel	100	79	71	71	64	64
GEV	100	92	92	100	92	92

TABLE 7 Linear and non-linear regression models for regionalization of maximum srteamflow for the Sapucaí River watershed

Equation	a	b	R^2	d	RQME ($m^3\ s^{-1}$)
$Q_{max5} = a + b \cdot Da$	27.482	0.113	0.986	0.996	23.665
$Q_{max5} = a \cdot P_{eq}^{b}$	27.482	2.437	0.986	0.996	23.665
$Q_{max5} = a \cdot P_{eq750}^{b}$	26.849	5.023	0.985	0.996	24.032
$Q_{max10} = a + b \cdot Da$	32.327	0.132	0.974	0.993	37.420
$Q_{max10} = a \cdot P_{eq}^{b}$	32.327	2.847	0.974	0.993	37.420
$Q_{max10} = a \cdot P_{eq750}^{b}$	31.611	5.866	0.974	0.993	37.910
$Q_{max20} = a + b \cdot Da$	37.456^{ns}	0.149	0.959	0.989	53.867
$Q_{max20} = a \cdot P_{eq}^{b}$	37.456^{ns}	3.217	0.959	0.989	53.867
$Q_{max20} = a \cdot P_{eq750}^{b}$	36.671^{ns}	6.629	0.958	0.989	54.439
$Q_{max50} = a + b \cdot Da$	45.228^{ns}	0.170	0.933	0.982	79.525
$Q_{max50} = a \cdot P_{eq}^{b}$	45.228^{ns}	3.667	0.933	0.982	79.525
$Q_{max50} = a \cdot P_{eq750}^{b}$	44.364^{ns}	7.555	0.932	0.982	80.169
$Q_{max100} = a + b \cdot Da$	52.156^{ns}	0.185	0.909	0.976	101.890
$Q_{max100} = a \cdot P_{eq}^{b}$	52.156^{ns}	3.984	0.909	0.976	101.890
$Q_{max100} = a \cdot P_{eq750}^{b}$	51.241^{ns}	8.207	0.908	0.975	102.572

$Q_{max5} = a \cdot Da^b$	0.351	0.873	0.985	0.996	24.960
$Q_{max5} = a \cdot P_{eq}^b$	5.117	0.873	0.985	0.996	24.960
$Q_{max5} = a \cdot P_{eq750}^b$	9.370	0.878	0.984	0.996	25.574
$Q_{max10} = a \cdot Da^b$	0.420^{ns}	0.870	0.974	0.993	38.251
$Q_{max10} = a \cdot P_{eq}^b$	6.071	0.870	0.974	0.993	38.251
$Q_{max10} = a \cdot P_{eq750}^b$	11.088	0.875	0.973	0.993	39.015
$Q_{max20} = a \cdot Da^b$	0.499^{ns}	0.864	0.959	0.989	54.241
$Q_{max20} = a \cdot P_{eq}^b$	7.083	0.864	0.959	0.989	54.241
$Q_{max20} = a \cdot P_{eq750}^b$	12.881	0.870	0.957	0.989	55.116
$Q_{max50} = a \cdot Da^b$	0.627^{ns}	0.853	0.934	0.983	79.280
$Q_{max50} = a \cdot P_{eq}^b$	8.603	0.853	0.934	0.983	79.280
$Q_{max50} = a \cdot P_{eq750}^b$	15.519	0.858	0.932	0.982	80.273
$Q_{max100} = a \cdot Da^b$	0.750^{ns}	0.842	0.911	0.976	101.140
$Q_{max100} = a \cdot P_{eq}^b$	9.946^{ns}	0.842	0.911	0.976	101.140
$Q_{max100} = a \cdot P_{eq750}^b$	17.806^{ns}	0.847	0.909	0.976	102.209

ns = não significativo pelo teste t de Student (H_0: a = 0; b=0)

TABLE 8 Linear and non-linear regression models for regionalization of average annual streamflow for the Sapucaí River watershed.

Equation	a	b	R^2	d	RQME (m^3 s^{-1})
$Q_{med} = a + b \cdot Da$	-0.434^{ns}	0.021	0.996	0.999	2.455
$Q_{med} = a \cdot b \cdot Peq$	-0.434^{ns}	0.455	0.996	0.999	2.455
$Q_{med} = a \cdot b \cdot Peq750$	-0.552^{ns}	0.937	0.995	0.999	2.589
$Q_{med} = a \cdot Da^b$	0.013	1.056	0.997	0.999	2.088
$Q_{med} = a \cdot P_{eq}{}^b$	0.336	1.056	0.997	0.999	2.088
$Q_{med} = a \cdot P_{eq750}{}^b$	0.697	1.063	0.997	0.999	2.157

TABLE 9 Linear and non-linear regression models for regionalization of average monthly flows for the Sapucaí River watershed

Equation	a	b	R^2	d	RQME (m^3 s^{-1})	Equation	a	b	R^2	d	RQME (m^3 s^{-1})
Qjan = a + b·Da	-2.292^{ns}	0.042	0.993	0.996	6.405	Qjan = a + b·Peq	-2.292^{ns}	0.912	0.993	0.998	6.405
Qfeb = a + b·Da	-2.722^{ns}	0.040	0.988	0.997	7.752	Qfeb = a + b·Peq	-2.722^{ns}	0.866	0.988	0.997	7.752
Qmar = a + b·Da	-1.157^{ns}	0.032	0.993	0.998	4.730	Qmar = a + b·Peq	-1.157^{ns}	0.683	0.993	0.998	4.730
Qapr = a + b·Da	-0.480^{ns}	0.023	0.995	0.999	2.978	Qapr = a + b·Peq	-0.480^{ns}	0.505	0.995	0.999	2.978
Qmay = a + b·Da	-0.082^{ns}	0.017	0.996	0.999	1.928	Qmay = a + b·Peq	-0.082^{ns}	0.376	0.996	0.999	1.928
Qjun = a + b·Da	0.093^{ns}	0.015	0.997	0.999	1.366	Qjun = a + b·Peq	0.093^{ns}	0.316	0.997	0.999	1.366
Qjul = a + b·Da	0.200^{ns}	0.012	0.997	0.999	1.128	Qjul = a + b·Peq	0.200^{ns}	0.267	0.997	0.999	1.128
Qaug = a + b·Da	0.146^{ns}	0.010	0.998	0.999	0.877	Qaug = a + b·Peq	0.146^{ns}	0.221	0.998	0.999	0.877
Qsep = a + b·Da	0.216^{ns}	0.010	0.997	0.999	0.929	Qsep = a + b·Peq	0.216^{ns}	0.226	0.997	0.999	0.929
Qoct = a + b·Da	0.129^{ns}	0.012	0.994	0.998	1.732	Qoct = a + b·Peq	0.129^{ns}	0.267	0.994	0.998	1.732
Qnov = a + b·Da	0.369^{ns}	0.015	0.997	0.999	1.339	Qnov = a + b·Peq	0.369^{ns}	0.327	0.99	0.999	1.339

									7		
$Qdec = a + b \cdot Da$	0.367[ns]	0.023	0.996	0.999	2.568	$Qdec = a + b \cdot Peq$	0.367[ns]	0.491	0.99 6	0.999	2.568
$Qjan = a \cdot Da^b$	0.021	1.081	0.995	0.999	5.445	$Qjan = a \cdot P_{eq}^b$	0.584	1.081	0.99 5	0.999	5.445
$Qfeb = a \cdot Da^b$	0.013	1.128	0.994	0.998	5.911	$Qfeb = a \cdot P_{eq}^b$	0.428	1.128	0.99 4	0.998	5.911
$Qmar = a \cdot Da^b$	0.016	1.083	0.996	0.999	3.835	$Qmar = a \cdot P_{eq}^b$	0.433	1.083	0.99 6	0.999	3.835
$Qapr = a \cdot Da^b$	0.014	1.060	0.997	0.999	2.545	$Qapr = a \cdot P_{eq}^b$	0.364	0.999	0.99 7	0.999	2.545
$Qmay = a \cdot Da^b$	0.012	1.040	0.997	0.999	1.753	$Qmay = a \cdot P_{eq}^b$	0.304	1.040	0.99 7	0.999	1.753
$Qjun = a \cdot Da^b$	0.012	1.024	0.998	0.999	1.305	$Qjun = a \cdot P_{eq}^b$	0.279	1.024	0.99 8	0.999	1.305
$Qjul = a \cdot Da^b$	0.011	1.017	0.998	0.999	1.111	$Qjul = a \cdot P_{eq}^b$	0.245	1.017	0.99 8	0.999	1.111
$Qaug = a \cdot Da^b$	0.009	1.012	0.998	0.999	0.872	$Qaug = a \cdot P_{eq}^b$	0.208	1.012	0.99 8	0.999	0.872
$Qsep = a \cdot Da^b$	0.010	1.009	0.998	0.999	0.937	$Qsep = a \cdot P_{eq}^b$	0.217	1.009	0.99 8	0.999	0.937
$Qoct = a \cdot Da^b$	0.009	1.040	0.995	0.999	1.640	$Qoct = a \cdot P_{eq}^b$	0.217	1.040	0.99 5	0.999	1.640

$Qnov = a \cdot Da^{b}$	0.014	1.007	0.998	0.999	1.364	$Qnov = a \cdot P_{eq}^{b}$	0.317	1.007	0.99 8	0.999	1.364
$Qdec = a \cdot Da^{b}$	0.023	1.000	0.996	0.999	2.584	$Qdec = a \cdot P_{eq}^{b}$	0.494	1.000	0.99 6	0.999	2.584

842

TABLE 70 Linear and non-linear regression models for regionalization of minimum flows for the Sapucaí River watershed

Equation	a	b	R^2	d	RQME ($m^3\ s^{-1}$)
$Q_{7,10} = a + b \cdot Da$	-0.307^{ns}	0.006	0.987	0.997	1.135
$Q_{7,10} = a + b \cdot P_{eq}$	-0.307^{ns}	0.124	0.987	0.997	1.135
$Q_{7,10} = a + b \cdot P_{eq750}$	-0.336^{ns}	0.256	0.986	0.997	1.185
$Q_{95} = a + b \cdot Da$	-0.106^{ns}	0.007	0.994	0.999	0.981
$Q_{95} = a + b \cdot P_{eq}$	-0.106^{ns}	0.158	0.994	0.999	0.981
$Q_{95} = a + b \cdot P_{eq750}$	-0.145^{ns}	0.326	0.993	0.998	1.052
$Q_{90} = a + b \cdot Da$	0.116^{ns}	0.008	0.996	0.999	0.864
$Q_{90} = a + b \cdot P_{eq}$	0.116^{ns}	0.179	0.996	0.999	0.864
$Q_{90} = a + b \cdot P_{eq750}$	0.072^{ns}	0.370	0.996	0.999	0.964
$Q_{80} = a + b \cdot Da$	0.211^{ns}	0.010	0.998	0.999	0.838
$Q_{80} = a + b \cdot P_{eq}$	0.211^{ns}	0.214	0.998	0.999	0.838
$Q_{80} = a + b \cdot P_{eq750}$	0.158^{ns}	0.442	0.997	0.999	0.961
$Q_{50} = a + b \cdot Da$	0.040^{ns}	0.015	0.998	0.999	1.226
$Q_{50} = a + b \cdot P_{eq}$	0.040^{ns}	0.330	0.998	0.999	1.226
$Q_{50} = a + b \cdot P_{eq750}$	-0.044^{ns}	0.680	0.997	0.999	1.387
$Q_{7,10} = a \cdot Da^b$	0.002	1.098	0.991	0.998	0.984
$Q_{7,10} = a \cdot P_{eq}^b$	0.072	1.098	0.991	0.998	0.984
$Q_{7,10} = a \cdot P_{eq750}^b$	0.154	1.106	0.991	0.997	1.018
$Q_{95} = a \cdot Da^b$	0.005	1.050	0.996	0.999	0.891
$Q_{95} = a \cdot P_{eq}^b$	0.121	1.050	0.996	0.999	0.891
$Q_{95} = a \cdot P_{eq750}^b$	0.249	1.057	0.995	0.999	0.944
$Q_{90} = a \cdot Da^b$	0.007	1.015	0.997	0.999	0.856
$Q_{90} = a \cdot P_{eq}^b$	0.166	1.015	0.997	0.999	0.856
$Q_{90} = a \cdot P_{eq750}^b$	0.335	1.022	0.996	0.999	0.941
$Q_{80} = a \cdot Da^b$	0.010	1.001	0.998	0.999	0.854
$Q_{80} = a \cdot P_{eq}^b$	0.215	1.001	0.998	0.999	0.854
$Q_{80} = a \cdot P_{eq750}^b$	0.428	1.008	0.997	0.999	0.964
$Q_{50} = a \cdot Da^b$	0.002	0.019	0.998	> 0.99	1.171
$Q_{50} = a \cdot P_{eq}^b$	0.297	1.020	0.998	> 0.99	1.171

$Q_{50} = a \cdot P_{eq750}^{b}$	0.600	1.027	0.998	0.999	1.299

FIGURE LEGENDS

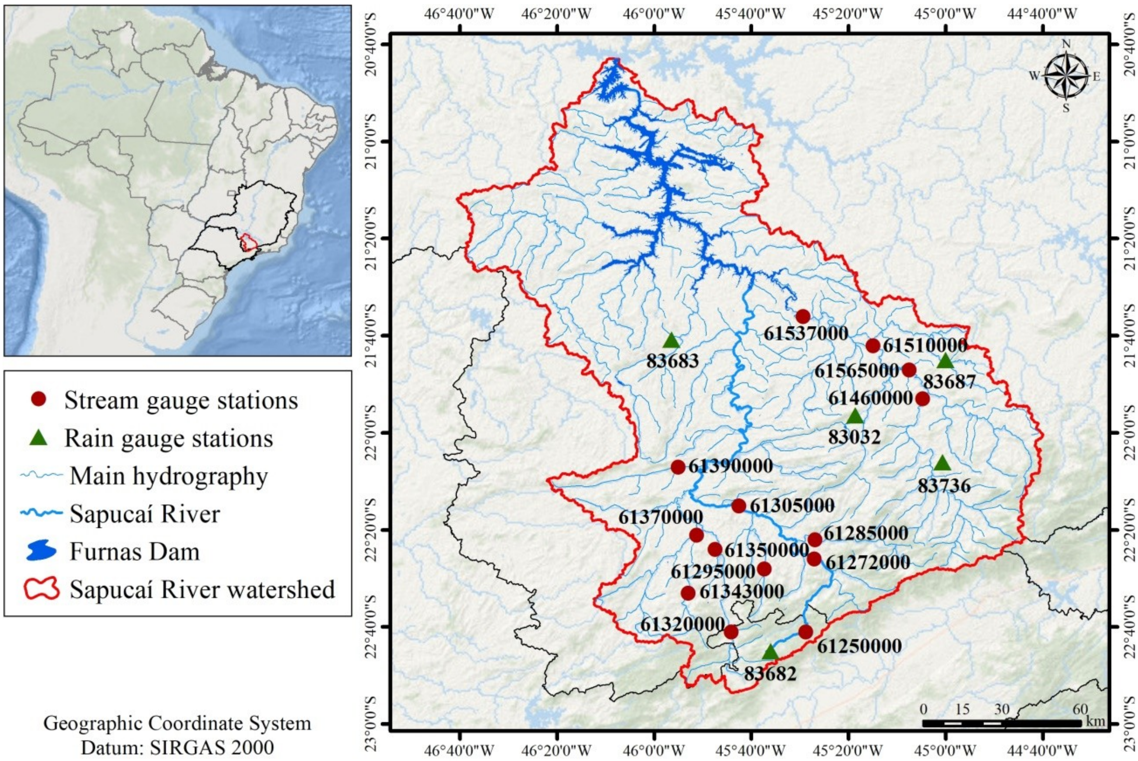

FIGURE 1 Sapucaí River watershed, highlighting the stream gauge stations and rain gauge stations used in the study

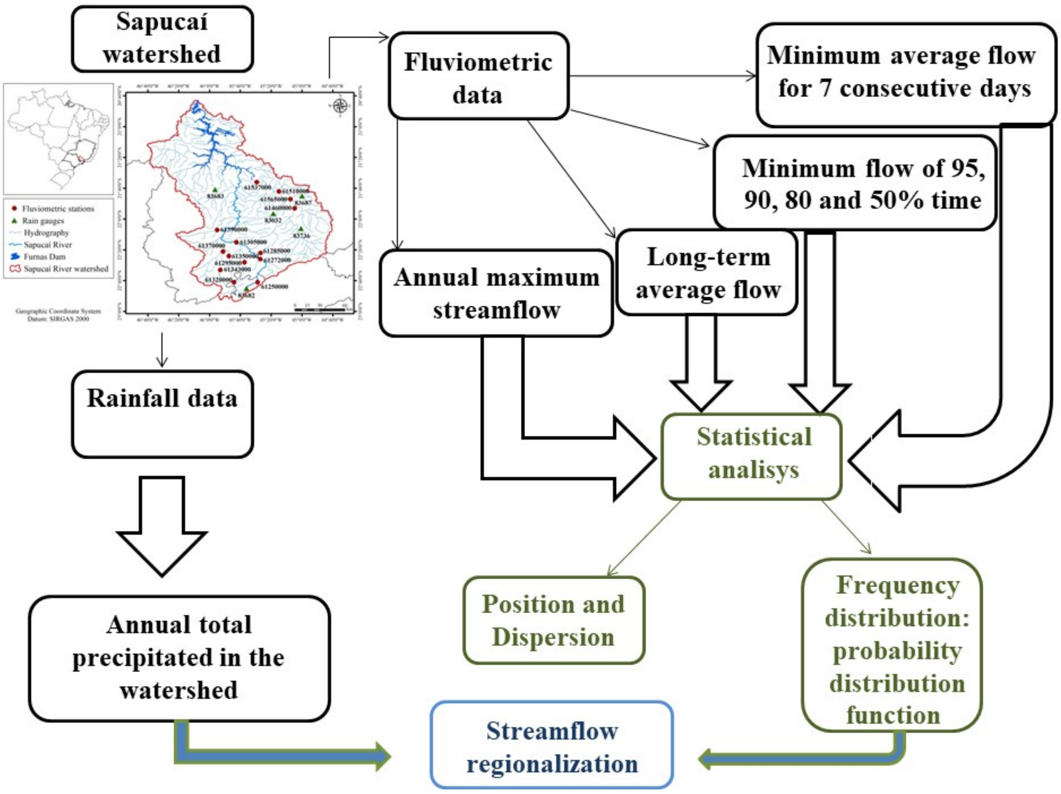

FIGURE 2 Flowchart of the methodology used

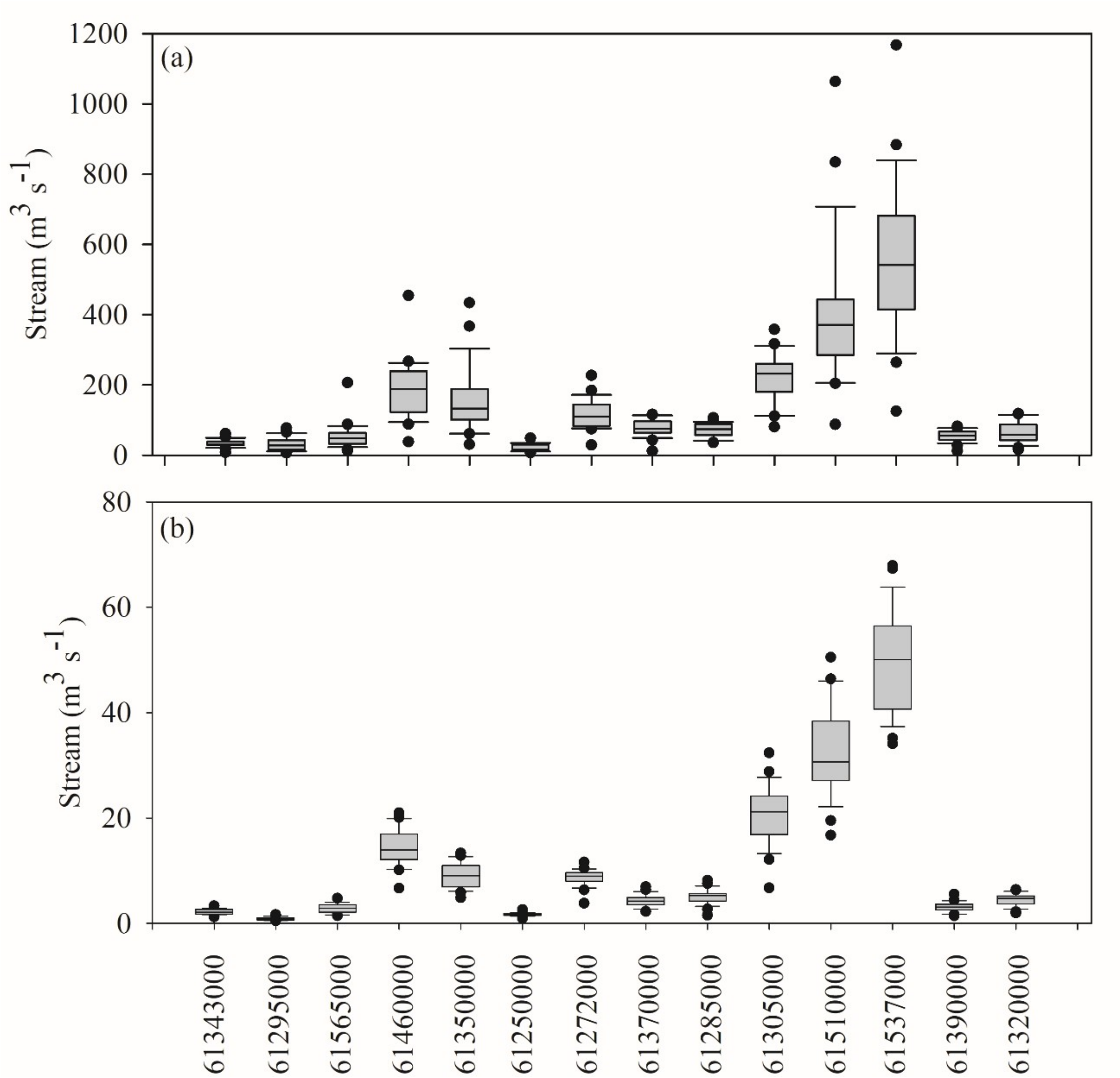

FIGURE 3 Box-plot of the maximum streamflow (b) and minimum streamflow of seven consecutive days (b) for the stream gauge stations in the Sapucaí River basin

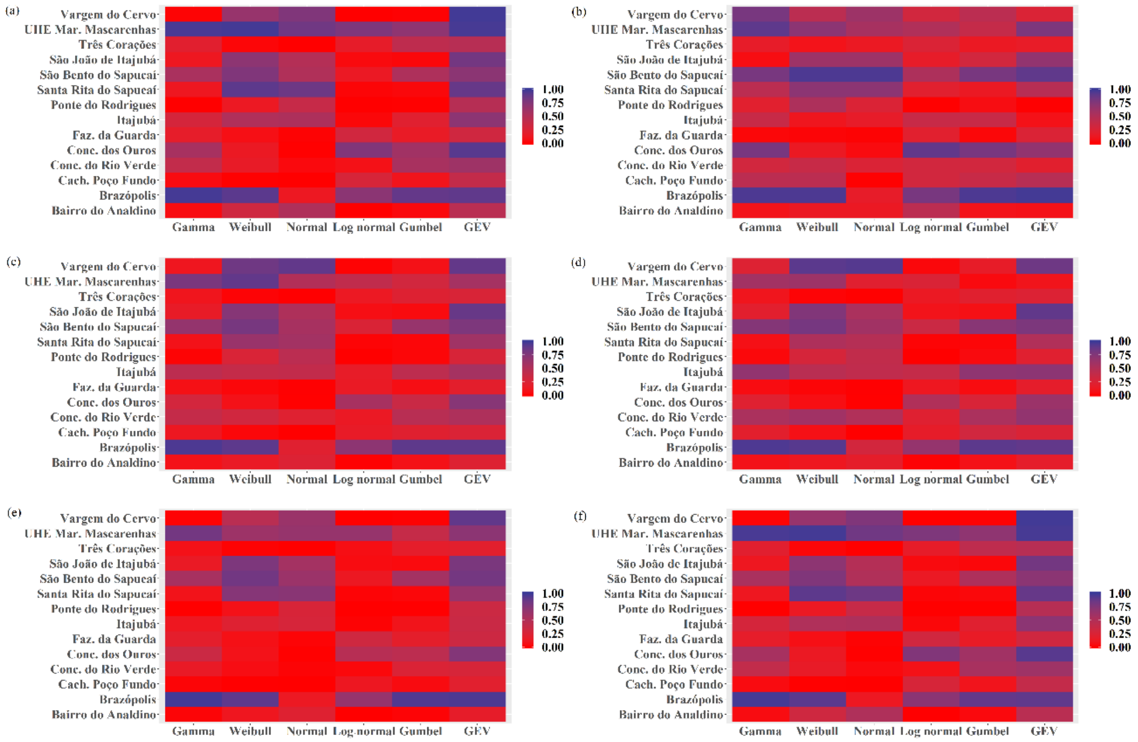

FIGURE 4 P-values classification obtained in the different goodness-of-fit tests to represent the maximum streamflow

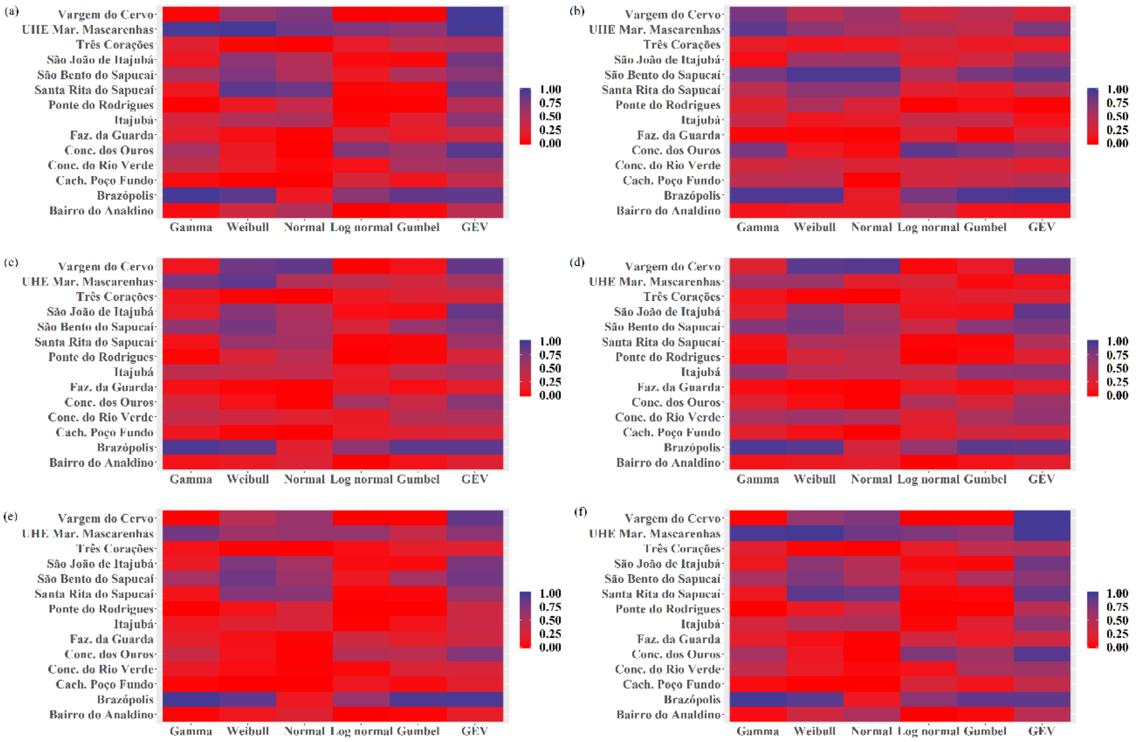

FIGURE 5 P-values classification obtained in the different goodness-of-fit tests to represent the minimum streamflow

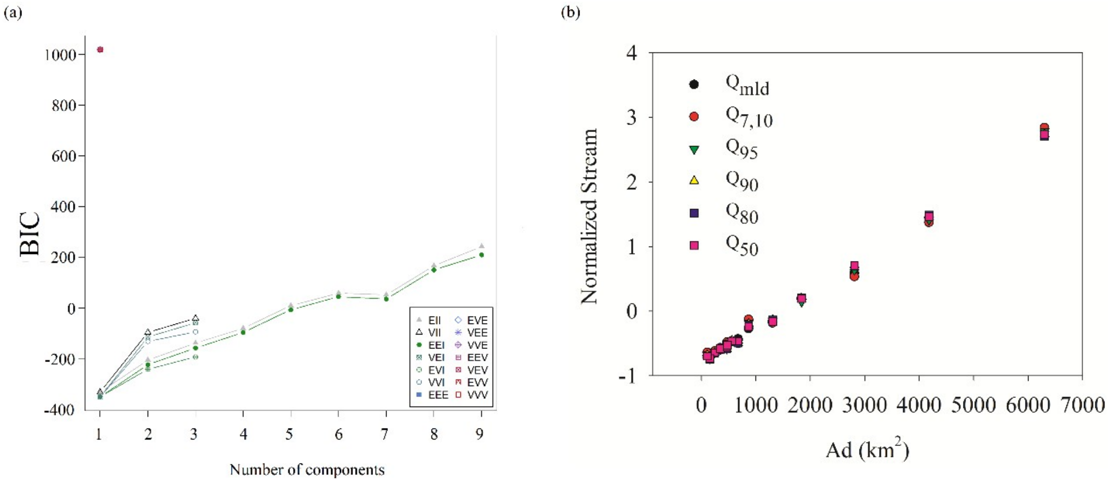

FIGURE 6 Bayesian Information Criterion (BIC) for choosing the agglomeration model with the optimal number of groups (a) and standard reference flows of each fluviometric station (b)

"EII" = spherical, equal volume ; "VII" = spherical, unequal volume; "EEI" = diagonal, equal volume and shape; "VEI" = diagonal, varying volume, equal shape; "EVI" = diagonal, equal volume, varying shape; "VVI" = diagonal, varying volume and shape; "EEE" = ellipsoidal, equal volume, shape, and orientation; "EVE" = ellipsoidal, equal volume and orientation; "VEE" = ellipsoidal, equal shape and orientation; "VVE" = ellipsoidal, equal orientation; "EEV" = ellipsoidal, equal volume and equal shape; "VEV" = ellipsoidal, equal shape; "EVV" = ellipsoidal, equal volume; "VVV" = ellipsoidal, varying volume, shape, and orientation.