Revista Brasileira de Recursos Hídricos
Brazilian Journal of Water Resources
Versão On-line ISSN 2318-0331
RBRH, Porto Alegre, v. 22, e27, 2017
Scientific/Technical Article

http://dx.doi.org/10.1590/2318-0331.021720160064

Hydrological regionalization of maximum stream flows using an approach based on L-moments

Regionalização hidrológica de vazões máximas utilizando uma abordagem baseada nos momentos-L

Felício Cassalho[1], Samuel Beskow[1], Marcelle Martins Vargas[1], Maíra Martim de Moura[1], Leo Fernandes Ávila[1] and Carlos Rogério de Mello[2]

[1]Universidade Federal de Pelotas, Pelotas, RS, Brazil
[2]Universidade Federal de Lavras, Lavras, MG, Brazil

E-mails: felicioufpel@gmail.com (FC), samuel.beskow@ufpel.edu.br (SB), marcellevarg@gmail.com (MMV), martimdemoura@gmail.com (MMM), avilalf@gmail.com (LFA), crmello@deg.ufla.br (CRM)

Received: December 15, 2016 - Revised: January 23, 2017 - Accepted: February 13, 2017

ABSTRACT

The proper design of hydraulic structures depends on estimates of maximum stream flows. The scarce stream flow monitoring in Brazil has led to the use of regionalization methods. The main objective of this study was to develop a tool via regional function to estimate maximum stream flows and their corresponding return periods (RP) with the aid of techniques based on the L-moments method, seeking for adequate hydrologic engineering applications and flood risk management. Annual maximum stream flow historical series were adjusted to traditional 2-parameter probability density functions (PDFs) (Normal, 2-parameter Log-Normal, Gumbel, Gamma) and multiparameter PDFs (GEV and Kappa), based on the L-moments method, which were used in the development of the regional function employing the dimensionless curve method. The regional function's predictive capability was determined by cross-validation for different RPs. It can be concluded that the approach based on L-moments was successfully used to adjust the regional function. In addition, the regional function: i) was improved when using the aforementioned multiparameter PDFs and ii) was framed as optimum for RP of up to 100 years and considered useful for practical engineering projects and flood risk management.

Keywords: Flood risk management; Statistical hydrology; GEV; Kappa; Mirim-São Gonçalo transboundary basin.

RESUMO

O dimensionamento adequado de estruturas hidráulicas é dependente das estimativas de vazões máximas. A escassez no monitoramento hidrológico no Brasil tem levado ao uso de métodos de regionalização. O principal objetivo do presente estudo foi desenvolver uma ferramenta via função regional para a estimativa das vazões máximas e seus respectivos tempos de retorno, com o suporte de técnicas baseadas no método dos momentos-L, com vistas a aplicações adequadas da engenharia hidrológica e da gestão de risco de cheias. As séries de vazão máxima diária anual foram ajustadas às funções densidade de probabilidade tradicionais de 2 parâmetros (Normal, 2-parameter Log-Normal, Gumbel, Gamma) e multiparâmetros (GEV and Kappa), baseando-se no método dos momentos-L, as quais foram utilizadas no desenvolvimento da função regional pelo método da curva adimensional. A capacidade preditiva da função regional foi determinada por validação cruzada para diferentes tempos de retorno. Pode-se concluir que a abordagem baseada nos momentos-L no ajuste da função regional foi satisfatória. Além disso,a função regional: i) foi aperfeiçoada quando as distribuições multiparâmetros acima mencionadas foram usadas e ii) foi classificada como ótima para tempos de retorno de até 100 anos sendo útil a projetos práticos de engenharia e na gestão de cheias.

Palavras-chave: Gestão de risco de cheia; Hidrologia estatística; GEV; Kappa; Bacia hidrográfica transfronteiriça Mirim-São Gonçalo.

INTRODUCTION

Appropriate estimates of the magnitude and frequency of maximum stream flow events are required for hydraulic structures projects, as well as for water resources management. Hydraulic structures such as bridges, dams, culverts, collection and drainage systems, among others, require information on design stream flows, which are related to maximum stream flow events associated with predefined return periods (RPs) (BESKOW et al., 2015; KUMAR et al., 2003). In water resources management, the understanding of the hydrological behavior of a given basin related to the flood events aids in mitigating the impact of natural disasters and with flooded-area management. Therefore, the proper estimate of the magnitude and frequency of maximum hydrological events is a limiting factor in the economic viability of both hydraulic structures projects and water resources management (MERZ et al., 2010).

Overestimated design stream flows may increase construction costs; on the other hand, the underestimation of these values indeed lead to poor dimensioning of hydraulic structures (SECKIN; HAKTANIR; YURTAL, 2011). Ideally, peak stream flows are estimated from historical series originated from the hydrological monitoring in the outlet of interest. However, Beskow et al. (2016a) point out that the unsatisfactory density of gauging stations operating in developing countries (e.g Brazil), associated with short time-series available, especially in small and medium-sized basins, hamper the analysis, thus leading to the use of indirect methods to estimate hydrologic variables. In this context, hydrological regionalization can be applied on gauged sites, in order to add information to the existing series, as well transfer them to ungauged locations for the sake of meeting the data demands (BESKOW et al., 2016b).

The stream flow series must be modeled probabilistically according to a suitable Probability Density Function (PDF), once they meet the stationarity requirement. The choice of the PDF is carried out based on goodness of fit tests, which are based on the comparison between symmetry and kurtosis of theoretical PDFs and empirical frequencies of the historical series (KUMAR; CHATTERJEE, 2005). Several authors have assessed the adjustment quality of different PDFs with respect to maximum stream flow series. Castellarin (2007) analyzed the performance of the Generalized Extreme Value (GEV) function, in the modeling of maximum stream flows in 33 basins, divided into three homogeneous regions in central and northern Italy. This researcher concluded that it was suitable for positive asymptotic series (positive tail), which is characteristic of the hydrological variable of interest. Guse, Hofherr and Merz (2010), following recommendations of Castellarin (2007), also found a good fit of the GEV in the study of maximum stream flows in the Saxony region, in eastern Germany. Eaton, Church and Ham (2002) sought among Normal, Log-Normal (LN), Gumbel, Pearson Type III (PT-III) and Log Pearson functions, for flooding studies in the state of British Columbia – Canada. They concluded that PT-III function was the most indicated for the region. Kjeldsen, Smithers and Schulze (2002), in a study in the state of Kwazulu-Natal, South Africa, concluded that the Generalized Pareto (GPA) function had a better performance when compared to other 4 PDFs applied to 29 historical series. In a study in the United States on flooding in the twentieth century, considering only outlets with over 100 years of monitoring, Villarini et al. (2009) found that LN and Gamma functions, both with 2 parameters, had performance superior than Gumbel and Weibull functions. However, it should be pointed out, as reported by Seckin, Haktanir and Yurtal (2011), that, in addition to the suitability of the PDFs, the regionalization of maximum stream flows involves the identification of hydrologically homogenous regions.

Hydrologically homogeneous regions defined based on geographical and administrative criteria tend not to be representative, resulting in overestimation of the hydrological variable in some areas and underestimation in others (GAUME et al., 2010). According to Hosking and Wallis (1997), to identify the heterogeneity of a region, the regional heterogeneity measure is necessary. This measure is based on the analysis of the statistical variability of the data series in the study area compared to what would be expected for a truly homogeneous region. Thus, also according to these researchers, a region is considered homogeneous if all the PDFs adjusted to the historical series are similar to each other (HOSKING; WALLIS, 1997). The measure of regional heterogeneity, H, presented by Hosking and Wallis (1997), has been widely used in scientific studies of hydrological regionalization involving different variables of interest (NOTO; LA LOGGIA, 2009; GAUME et al., 2010; SECKIN; HAKTANIR; YURTAL, 2011; BESKOW et al., 2016b).

Various techniques have been used in stream flow regionalization studies with the purpose of delimitating regions. The Region of Influence (ROI) (ZRINJI; BURN, 1994) provides that the area of influence of a given gauging station consists of stations with smaller discordancy D_i than the predetermined threshold value (BURN, 1997). In Canonical Correlation Analysis (CCA), according to Cavadias et al. (2001), initially, two types of data are defined: the hydrologic variables and geomorphological and meteorological characteristics of the basins. Basins, gauged or not, are then represented as points in space. Regionalization is carried out according to the similarity of the arrangement of these points (CAVADIAS et al., 2001). Also noteworthy are the clustering techniques, which consist of various multivariate statistical procedures for grouping the data into clusters that may or may not be overlapped (RAO; SRINIVAS, 2008). Ward's Hierarchical Clustering method, which aims to minimize the Euclidean distance of the local characteristics of each cluster, is one of the most common methods of area delimitation in stream flow regionalization studies (MALEKINEZHAD; NACHTNEBEL; KLIK, 2011). Khatibi et al. (2011) point out the increasing importance of artificial intelligence (AI) techniques in stream flow studies in the last 20 years. Among AI techniques, Beskow et al. (2016b) highlight the algorithms based on fuzzy logic, K-means, Partitioning Around Medoids, K-harmonic means and Genetic K-Means. Despite the different approaches, the presented methods have in common the need to establish homogeneous regions, thus allowing the development of representative regional functions (BESKOW et al., 2016b).

Several authors have addressed the regionalization of maximum stream flows through techniques that differ both in the data processing and in the methodology used to obtain the regional function (TUCCI, 2002; CASTELLARIN, 2007; NOTO; LA LOGGIA, 2009; MALEKINEZHAD; NACHTNEBEL;

KLIK, 2011; NAGHETTINI, 2017). Among the methods for the development of regional function, Tucci (2002) highlights: the selected values method, the regionalization of probability distribution parameters and the dimensionless curve method.

For the direct regression method, stream flow values are determined for fixed RPs, e.g., Q_{10}, Q_{20}, Q_{50}, Q_{100}, derived from the best adjusted PDF for each time series, having as a result regional regression functions for each predetermined RP. In the regionalization of PDF's parameters, it is necessary to choose a representative probabilistic model for all series contained in the hydrologically homogeneous region. Once the PDF with better performance, based on goodness-of-fit statistical tests, is determined, the regionalization of its parameters is conducted, yielding a single function applicable for different RPs. Finally, there is the dimensionless curve method, which comprises two parts: the first determines the dimensionless curve and the second gives the average maximum stream flow function. In this method, it is possible to consider peak stream flows of different magnitudes since the stream flow values are non-dimensionalised for each historical series. It is important to note that regardless of the method chosen, it is necessary to choose explanatory variables for the study area that are capable of reducing the uncertainty of the estimates. Also, it is advisable to conduct the validation of the resulting regional function, thus making it possible to quantify the reliability of the estimates (CASTELLARIN, 2007). In this study, the proposed methodological sequence was applied to Mirim-São Gonçalo basin.

The main objective of this study was to develop an approach primarily based on the L-moments to generate a regional function tool intended for estimating maximum stream flows associated with return periods of interest in the Mirim-São Gonçalo transboundary basin, and evaluate its predictive capability with a view towards hydrologic and hydraulic projects as well as flood risk management. Moreover, this study aimed to verify if multiparameter PDFs have advantages over traditional 2-parameter PDFs with respect to adjustment of the maximum stream flows series considered for hydrological regionalization in this study.

MATERIAL AND METHODS

The Mirim-São Gonçalo basin has an area of approximately 62,250 km², located between 31º30'S and 34º30'S latitudes and 52ºW and 56ºW longitudes; 47% in Brazilian territory (Figure 1) and 53% in Uruguay (CORADI et al., 2009). According to historical series of annual rainfall from 1961 to 2016, provided by the National Institute of Meteorology for the municipalities whose territory is fully or partially covered by the Mirim-São Gonçalo basin, the average annual total rainfall ranges from 1,225 mm in the city of Rio Grande to 1,507 mm, in the city of Bagé (INMET, 2016). The climate is classified as humid subtropical (Cfa), according to the Koppen classification method (SPAROVEK; LIER; DOURADO NETO, 2007).

Of the 85 gauging stations available in the HidroWeb portal of the National Water Agency (ANA) for the Mirim-São Gonçalo basin, only 7 had historical series formed by daily stream flow data sets that met the demands of this study. From these historical series were structured series of annual maximum stream flow (Figure 1 and Table 1).

The importance of analyzing the stationarity of the series for the study of regionalization should be emphasized. According to Salas (1992), a hydrological series is stationary if it does not have trends, shifts or periodicity. Therefore, from a hydrological point of view, trend and stationarity are directly related, which means that the trend analysis of a given historical series allows inferring about the stationarity. For these series, the nonparametric Mann-Kendall (MK) test was applied, as described by Wang et al. (2015), in order to verify existence of trends. The null hypothesis of MK test (H_0) states there is no trend in the data over the time, since they are samples of *n* random variables, independent and equally distributed (ISHAK et al., 2013).

In order to verify if the studied region is hydrologically homogeneous, the regional heterogeneity measure (H) (HOSKING; WALLIS, 1997), which takes into account the magnitude of L-moments of the samples in relation to what is expected for a truly homogeneous region, was applied. Once regional homogeneity is demonstrated, it is possible to transfer information from hydrologically similar sites to the outlet of interest (BURN, 1997). The measure of discordancy (D_i) proposed by Hosking and Wallis (1995), can be used in a complementary way to identify series whose L-moments quotients are significantly different when compared to other series. This process is conducted in two steps (NOTO; LA LOGGIA, 2009). Firstly, D_i is applied to a large number of stations distributed over a wide area. Series with considerable errors in relation to others are classified as discordant and should be analyzed regarding their reliability in data collection and processing. Secondly, with the study area already defined, the discordancy for each series in the analyzed region is calculated. If any series is discordant, its removal and possible regrouping in an adjacent region should be considered. The proven non-discordant series can then be modeled probabilistically and used in the development of the regional function.

Normal, 2-parameter Log-Normal (LN-2P), Gumbel, Gamma, Kappa and GEV PDFs, as described by Hosking and Wallis (1997), were adjusted to the annual maximum stream flow historical series. Despite its simplicity and widespread use, the method of moments performance is often unsatisfactory, especially for shorter historical series and probability distributions with three or more parameters, resulting in inaccurate estimates of the PDF parameters (VIVEKANANDAN; SHUKLA, 2015). Thus, the parameters of the aforementioned probabilistic models were obtained according to the L-moments method, which has better

Table 1. Sub-basins characteristics.

Gauging station	Drainage area (km²)	Length of the historical series (years)
Passo do Ricardo	5,416.3	18
Pedro Osório	4,699.5	12
Picada Nova	2,238.5	9
Cerro Chato	1,047.6	27
Passo dos Carros	131.2	46
Ponte do Império	1,839.6	9
Ponte Cord. de F.	368.6	47

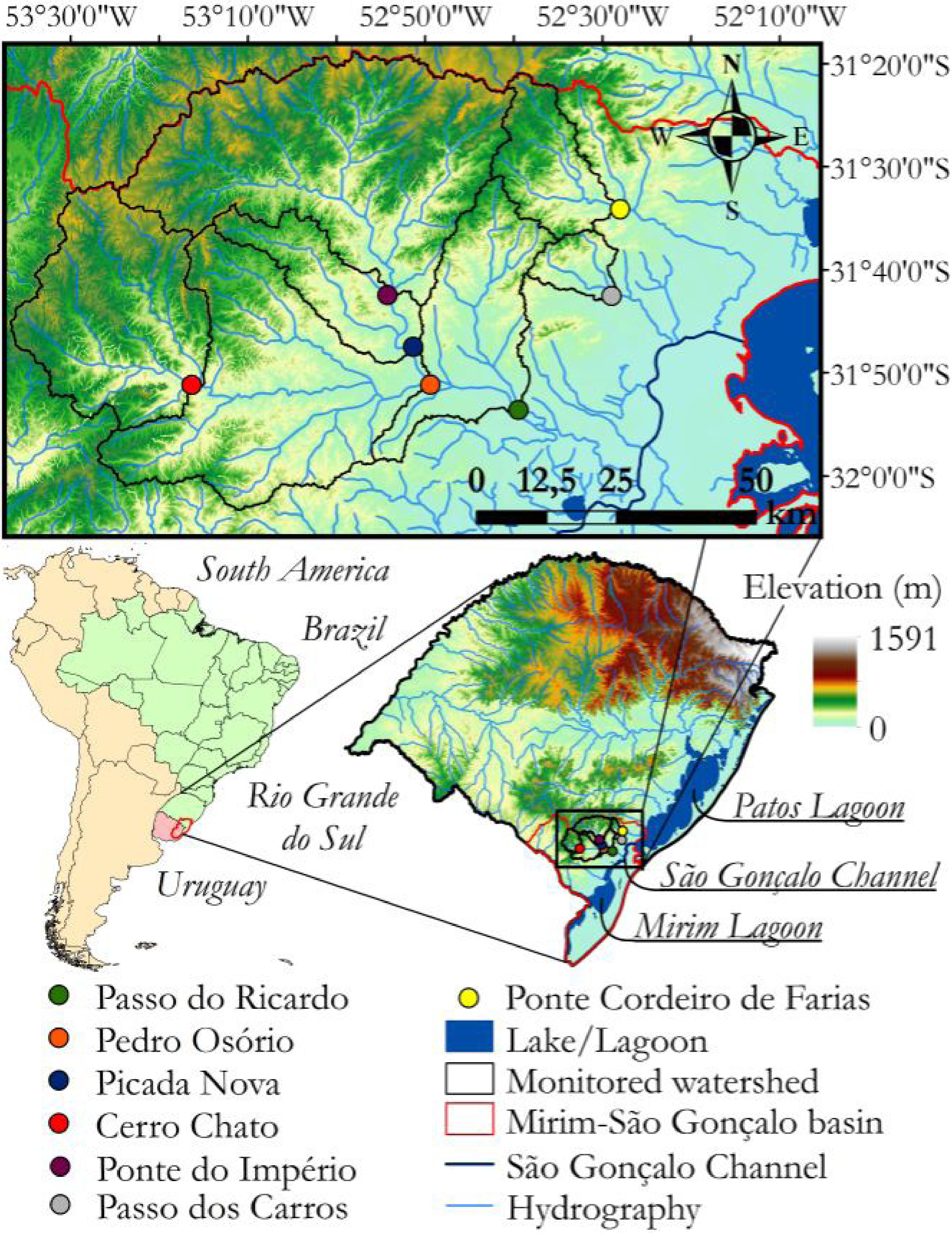

Figure 1. Geographical location of Mirim-São Gonçalo basin, considering its area in Uruguay and in Brazil (Rio Grande do Sul State), and the sub-basins evaluated in this study.

statistical properties, leading to a better efficacy of the estimators (KUMAR et al., 2003).

Because maximum stream flow tends to follow an asymptotic distribution, a better fit of the distribution tails is necessary (BESKOW et al., 2015). In this context, the Anderson Darling test (AD), which is considered superior (L'ECUYER; SIMARD, 2007) and more appropriate for this situation, as it assigns a greater importance to the extremes of the cumulative probability functions (HEO et al., 2013), was applied to evaluate the performance of the PDFs.

The preparation of the annual maximum stream flow series, the calculation of H and D_i measures, the adjustment of probabilistic models and goodness-of-fit test were carried out with the aid of the computer program "System of Hydrological Data Acquisition and Analysis" (SYHDA) (BESKOW et al., 2013). The identification of the study site, and the determination of drainage the area, as an explanatory variable used in the development of the regional function, were performed with the aid of ArcMap 10.4.1 software, using a digital elevation model derived from cartographic vector base of the Rio Grande do Sul State (HASENACK; WEBER, 2010).

The regional function was determined according to the dimensionless curve method, as suggested by Tucci (2002). The Q/Q_{mc} values correspond to the dimensionless annual maximum stream flow series in relation to the average maximum stream flow (Q_{mc}). Q_{mc} function, in turn, is determined from a regression equation relating it to explanatory variables, e.g., area, average annual rainfall, length and steepness of main river, drainage density, among others. Equation 1 represents the function evaluated to express the dimensionless curve, which has the following adjustment parameters: a, b, a' and b', A is the area (km^2) and RP the return period (years). The regionalized stream flow (Q_R) is given in m^3/s.

$$Q_R = \frac{Q}{Q_{mc}} \times Q_{mc} = [a \times \ln(RP) + b] \times (a' \times A^{b'}) \qquad (1)$$

After the verification of the PDFs' goodness of fit by the AD test, the performance of the generated regional function was evaluated, as well as its predictive capability. The predictive capability was evaluated by the cross-validation method, as proposed by Vezza et al. (2010), considering the RPs of 10, 20, 50 and 100 years. To compare the observed values, given by the estimated quantiles, to those estimated from the cross-validation process, the performance statistic "c" was employed (CAMARGO; SENTELHAS, 1997). The values of "c" statistic were then classified according to recommendations of Camargo and Sentelhas (1997): Optimum, c > 0.85; very good, 0.76 to 0.85; good, 0.66 to 0.75; median, 61 to.65; tolerable, 0.51 to 0.60; bad, 0.41 to 0.50 and terrible c ≤ 0.40.

RESULTS AND DISCUSSION

The 7 historical series considered (Figure 1) presented no trend, according to the MK test at a 5% significance level (Table 2); thus, all the historical series were used for the following steps of this study. As described by Makridakis, Wheelwright and Hyndman (1998), in hydrologically stationary series, the magnitude of the values found for the hydrological variable of interest are distributed around the average, because the processes that influence the hydrological variable of interest are in balance. Moreover, the variance is constant over time, reflecting no temporal trend in the data. The existence of trends in stream flow series occurs due to anthropogenic causes, e.g., changes in the land use of the basin and/or the gauging outlet, as well as to natural causes resulting from the biogeochemical cycles of the planet, e.g., ENSO (ISHAK et al., 2013).

Table 2. Statistics of the historical series corresponding to the gauging stations considered.

Gauging station	MK*	D	PDF**	AD
Passo do Ricardo	0.2	0.96	GEV	0.31
Pedro Osório	0.64	1.13	GEV	0.13
Picada Nova	0.07	1.92	Normal	0.45
Cerro Chato	0.93	0.42	Kappa	0.21
Passo dos Carros	0.13	0.69	GEV	0.15
Ponte do Império	0.61	1.42	GEV	0.41
Ponte Cord. de F.	0.80	0.46	Kappa	0.18

* Mann-Kendall for p > 0.05. ** Probability Density Function (PDF) that presented the best fit according to the Anderson-Darling (AD) test.

Inferences and statistical analysis in hydrological series are strongly based on the stationarity hypothesis, since the non-stationarity requires the temporal change in the parameters of the adjusted models, making analysis troublesome (ZHANG et al., 2014). The MK test has often been used in hydrological studies, demonstrating its potential to understand the temporal behavior of historical series aimed at hydrologic modeling (ZHANG et al., 2014; WANG et al., 2015; BESKOW et al., 2016b).

The region, consisting of the seven sub-basins (Figure 1), according to the historical series of the respective outlets, can be considered homogeneous, since the measure of regional heterogeneity, H (HOSKING; WALLIS, 1997), resulted in H = 0.02 (H<1) due to the low dispersion among the L-moments ratios of the gauged stations. Complementarily to the regional heterogeneity measure H, are presented in Table 2 the results regarding the discordancy measure, D_i (HOSKING; WALLIS, 1995). One can note in Table 2 that none of the historical series presented a D_i statistical value greater than the critical value suggested by Hosking and Wallis (1995), as being equal to 1.92 for seven gauging stations. It should be emphasized that Hosking and Wallis (1995) recommend the implementation of the discordancy measure for at least 7 historical series, due to proximity of the D_i critical values and the algebraic limits of the statistic. Therefore, the seven analyzed historical series are not in discordance with each other, thus presenting similar statistical characteristics. According to Neykov et al. (2007), high D_i values for a given historical series suggest: a) further investigation to identify the presence of errors in the data; or b) analysis of the possibility of moving to another region. The D_i measure has the potential to guide the gauging station removal process in regions classified as heterogeneous, as described by Beskow et al. (2016b) in a study of hydrological regionalization of Q_{90} in the state of Rio Grande do Sul, Brazil. From the results presented in Table 2, it can be concluded that there is no need to remove any of the gauging stations for the hydrological regionalization study.

According to the Anderson-Darling non-parametric test (AD), the GEV model was the best PDF for most of the annual maximum stream flow series (Table 2). This non-parametric test was specifically used for being quite robust and restrictive in identifying probabilistic models appropriate to represent asymptotic series, agreeing with the findings of other scientific studies (BEN-ZVI, 2009; FRANCO et al., 2014; BESKOW et al., 2015).

The good applicability of the GEV model, which has shape (κ), scale (α) and position (β) parameters, was also observed by Seckin, Haktanir and Yurtal (2011) as to the analysis of maximum stream flows in Turkey, where the model was the best adjusted to 498 of the 543 gauging stations considered. Malekinezhad, Nachtnebel and Klik (2011), comparing five 3-parameter PDFs, found that the GEV model was the only suitable for all the 3 homogeneous regions identified in the Namak-Lake basin, central Iran. Noto and La Loggia (2009), in a case study on the island of Sicily - Italy, pointed the GEV model as the most robust for all delineated homogeneous regions when compared to the other 3 PDFs. Kumar and Chatterjee (2005) also highlight the robustness of the GEV model in probabilistic modeling of maximum stream flows in a study applied to the northern basin of the Brahmaputra river - India. It is noteworthy that the above studies were also based on the L-moments method.

The other historical series were adjusted by Kappa and Normal probabilistic models. The Kappa, with parameters scale α, position ξ, and shape κ and h, was the best fitted, according to the AD test, for two historical series. The superior applicability of Kappa, over the GEV, Gumbel and LN-2P PDFs, was observed by Beskow et al. (2015) when analyzing the annual maximum daily rainfall in the state of Rio Grande do Sul, Brazil. Beskow et al. (2015) were one of the pioneers in the use of the Kappa PDF applied to rainfall series in Brazil. It is worthwhile to mention that the present study is one of the first in applying Kappa distribution in a maximum stream flow frequency analysis in the country.

Finally, the Normal model was the best fit for one historical series. The Normal model, with scale (mean - μ) and position (standard deviation - σ) parameters, is recommended to represent random variables fluctuating symmetrically around the mean and is indicated for series resulting from the sum of many other independent variables, which does not correspond the maximum stream flow series (NAGHETTINI, 2017), since they tend to be asymptotically positive. However, the small size of the series from the Picada Nova station may have influenced the adjustment of the Normal distribution, since its short monitoring period may not have captured the asymptotic behavior of the probability distribution. The inferiority of the Normal PDF and other traditional 2-parameter PDFs evaluated in this study for the adjustment of maximum stream flow series, when compared to multiparameter PDFs, has been extensively reported in the literature (EATON; CHURCH; HAM, 2002; KUMAR et al., 2003; KUMAR; CHATTERJEE, 2005).

None of the 7 historical series was better fitted by the LN-2P, Gumbel or Gamma PDFs (Table 2). Despite the fact that these models are recommended for maximum historical series once they reproduce its positive asymmetry, because they have only 2 parameters, they tend to present a lower performance than the multiparameter PDFs, i.e. GEV and Kappa. The improved estimate of maximum hydrological variables resulting from the application of multiparameter PDFs, compared to traditional two-parameter ones, i.e., Normal, LN-2P, Gumbel and Gamma, is widely demonstrated in studies conducted in various regions of the world (EATON; CHURCH; HAM, 2002; KJELDSEN; SMITHERS; SCHULZE, 2002; KUMAR et al., 2003; KUMAR; CHATTERJEE, 2005; NOTO; LA LOGGIA, 2009; SECKIN; HAKTANIR; YURTAL, 2011; BESKOW et al., 2015). The evaluation of traditional and multiparameter PDFs for the adjustment of annual maximum stream flow series as performed in this study, corroborates the findings of Beskow et al. (2015), who reported on the importance of the analysis of various theoretical PDFs, from the simplest to those with multiparameters, in order to obtain that which best represents the frequency distribution of the sample.

Using the drainage area as an explanatory variable for each of the 7 sub-basins analyzed, it was possible to develop a regression function for Q_{mc}, following the potential model as shown in Equation 1. The parameters a' and b', were equal to 4.957 and 0.696, respectively, resulting in a function with a performance index c of 0.99, classified as optimum, as proposed by Camargo and Sentelhas (1997) (Figure 2). Regarding the hydrological regionalization, the number of gauging stations used to represent a region is always a matter of concern, especially given the limited amount of historical series for analysis in the area of interest. In this study, historical series of 7 gauging stations were employed, exceeding the minimum number (6 stations) recommended by Ribeiro, Marques and Silva (2005), who substantiate this threshold based on the evaluation of homogeneous regions in the Doce river basin (Brazil).

For the Q/Q_{mc} function, the dimensionless stream flows were correlated with their RPs. Parameters a = 0.466 and b = 0.551, with a performance index c of 0.91, were obtained for this function, also categorized as optimum (CAMARGO; SENTELHAS, 1997) (Figure 3).

The product of these two equations culminated in the regional function (Equation 2) by the dimensionless curve method, as described by Tucci (2002).

$$Q_R = [0.466 \times \ln(RP) + 0.551] \times (4.957 A^{0.696}) \tag{2}$$

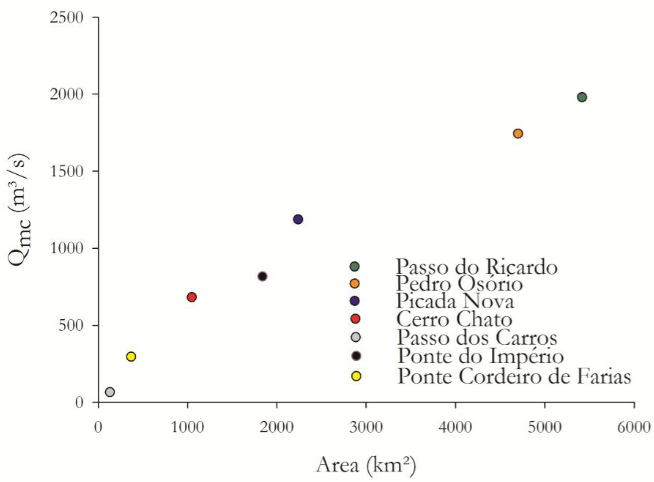

Figure 2. Relationship between average maximum stream flow and drainage area considering the basins used in this study.

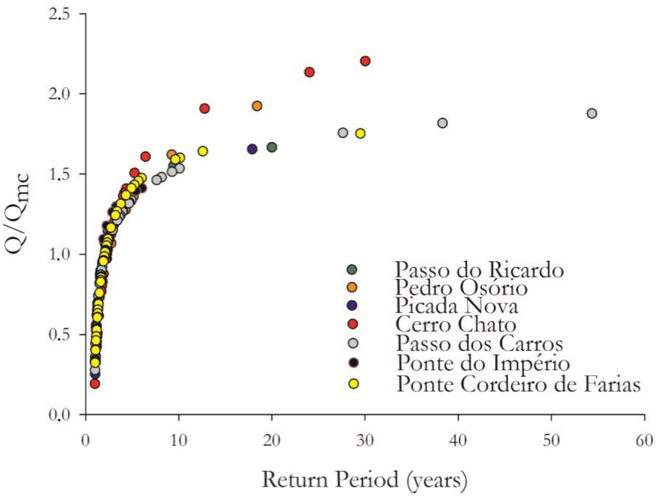

Figure 3. Regional function relating dimensionless stream flows to the corresponding RPs taking into account the basins used in this study.

Despite its easy applicability and simplicity regarding the determination of the explanatory variables, i.e. area and RP, the regional function has limitations that need to be stressed out. Since the areas of the analyzed basins vary from 131.2 to 5,416.3 square kilometers, estimates of maximum stream flows in river basins that extrapolate these limits in area may result in inaccurate estimates, as evidenced by Hosking and Wallis (1997), Kumar and Chatterjee (2005), and Naghettini (2017).

The predictive capability of the regional function, which was assessed by cross-validation according to the methodology proposed by Vezza et al. (2010), taking as reference the RPs of 10, 20, 50 and 100 years, demonstrates the excellent applicability of the function, even with the increased RPs (Table 3). The evaluation of the regional function's predictive capability in the studied region, by means of cross-validation, is of critical importance. According to Vezza et al. (2010), this method of cross-validation has advantages (e.g. robustness and applicability to all regionalization models) over other techniques used to assess predictive errors.

The predictive capability was found to be optimum for all the proposed RPs (Table 3). The authors consider that among the reasons for optimal function performance is the strong correlation observed between the drainage area and maximum stream flow, being able to infer that just the drainage area accounts for more than 90% of the maximum stream flow behavior for the most critical RP. In addition, the use of only the drainage area as explanatory variable is consistent with the principle of parsimony, which seeks to explain the phenomenon analyzed with the lowest possible number of explanatory variables. However, it is not suggested to use the regional function for RPs greater than 2 times the size of the historical series, since estimation of quantiles under these conditions is not recommended (NAGHETTINI, 2017). Although the drainage area has been an appropriate explanatory variable, the use of other variables correlated with maximum stream flows described in the literature (HOSKING; WALLIS, 1997) could increase the correlation between the estimated and observed values. Noto and La Loggia (2009) also used drainage area as an explanatory variable in their study; however, they had a gradual

Table 3. Correlation (r), accuracy (d) and performance (c) of the regional equation for different RPs.

RP (years)	10	20	50	100
r	0.98	0.97	0.96	0.95
d	0.99	0.99	0.98	0.97
c	0.97	0.96	0.94	0.92
Performance	Optimum	Optimum	Optimum	Optimum

increase in the regional function's performance with the addition of other parameters, such as, main river length, average rainfall for a 2-year RP and duration of 12 and 24 hours. Malekinezhad, Nachtnebel and Klik (2011) also pointed out a higher correlation between maximum stream flows and the main river length, thereby emphasizing the importance of the correlation analysis during the process of choosing the explanatory variables.

CONCLUSIONS

Based on the results obtained in this study, it can be concluded that the regional function was successfully developed based on L-moments techniques. Moreover, the regional function: i) has considerable improvements when using multiparameter PDFs rather than traditional 2-parameter PDFs; ii) presents suitable adjustment according to statistics used for a proven homogeneous region; iii) has adequate predictive capability, substantiated in the cross-validation procedure; and iv) is simple to use, requiring only one explanatory variable, being easily applied to practical engineering projects and flood risk management. Finally, future studies may consider a larger number of explanatory variables in order to improve maximum stream flow estimates.

ACKNOWLEDGEMENTS

The authors wish to thank CNPq for scholarships to the second (307523/2014-4) and sixth (303059/2013-3) authors, to UFPel for scholarship to the third author, to CAPES for scholarships to the fourth and fifth authors, to FAPEMIG for research grant (PPM VIII 071/2014) to the sixth author, to CNPq (485279/2013-4) and FAPERGS (2082-2551/13-0) for research grants to the second author.

REFERENCES

BEN-ZVI, A. Rainfall intensity–duration–frequency relationships derived from large partial duration series. *Journal of Hydrology*, v. 367, n. 1-2, p. 104-114, 2009. http://dx.doi.org/10.1016/j.jhydrol.2009.01.007.

BESKOW, S.; CALDEIRA, T. L.; MELLO, C. R.; FARIA, L. C.; GUEDES, H. A. S. Multiparameter probability distributions for heavy rainfall modeling in extreme southern Brazil. *Journal of Hydrology: Regional Studies*, v. 4, p. 123-133, 2015. http://dx.doi.org/10.1016/j.ejrh.2015.06.007

BESKOW, S.; CORRÊA, L. L.; MAHL, M.; SIMÕES, M. C.; CALDEIRA, T. L.; NUNES, G. S.; LUCAS, E. H.; FARIA, L. C.; MELLO, C. R. Desenvolvimento de um sistema computacional de aquisição e análise de dados hidrológicos. In: SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS, 20., 2013, Bento Gonçalves, RS. *Anais*... Porto Alegre: ABRH, 2013. CD-ROM.

BESKOW, S.; MELLO, C. R.; VARGAS, M. M.; CALDEIRA, T. L.; DURÃES, M. F.; AGUIAR, M. S. Artificial intelligence techniques coupled with seasonality measures for hydrological regionalization of Q90 under Brazilian conditions. *Journal of Hydrology*, v. 541, p. 1406-1419, 2016b. http://dx.doi.org/10.1016/j.jhydrol.2016.08.046.

BESKOW, S.; TIMM, L. C.; TAVARES, V. E. Q.; CALDEIRA, T. L.; AQUINO, L. S. Potential of the LASH model for water resources management in data-scarce basins: a case study of Fragata River Basin, Southern Brazil. *Hydrological Sciences Journal*, v. 61, n. 14, p. 2567-2578, 2016a. http://dx.doi.org/10.1080/02626667.2015.1133912.

BURN, D. H. Catchment similarity for regional flood frequency analysis using seasonality measures. *Journal of Hydrology*, v. 202, n. 1-4, p. 212-230, 1997. http://dx.doi.org/10.1016/S0022-1694(97)00068-1.

CAMARGO, A. P.; SENTELHAS, P. C. Avaliação de desempenho de diferentes métodos de estimativa da evapotranspiração potencial no estado de São Paulo, Brasil. *Revista Brasileira de Agrometeorologia*, v. 5, n. 1, p. 89-97, 1997.

CASTELLARIN, A. Probabilistic envelope curves for design flood estimation at ungauged sites. *Water Resources Research*, v. 43, n. 4, p. 1-12, 2007. PMid:20300476. http://dx.doi.org/10.1029/2005WR004384.

CAVADIAS, G. S.; OUARDA, T. B. M. J.; BOBEÉ, B.; GIRARD, C. A canonical correlation approach to the determination of homogeneous regions for regional flood estimation of ungauged basins. *Hydrological Sciences Journal*, v. 46, n. 4, p. 499-512, 2001. http://dx.doi.org/10.1080/02626660109492846.

CORADI, P. C.; PEREIRA-RAMIREZ, O.; FIA, R.; MATOS, A. T. Qualidade de água superficial da Bacia Hidrográfica da Lagoa Mirim. *Revista de Ciências Ambientais*, v. 3, p. 53-64, 2009.

EATON, B.; CHURCH, M.; HAM, D. Scaling and regionalization of flood flows in British Columbia, Canada. *Hydrological Processes*, v. 16, n. 16, p. 3245-3263, 2002. http://dx.doi.org/10.1002/hyp.1100.

FRANCO, C. S.; MARQUES, R. F. P. V.; OLIVEIRA, A. S.; OLIVEIRA, L. F. C. Distribuição de probabilidades para precipitação máxima diária na Bacia Hidrográfica do Rio Verde, Minas Gerais. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v. 18, n. 7, p. 735-741, 2014. http://dx.doi.org/10.1590/S1415-43662014000700010.

GAUME, E.; GAÁL, L.; VIGLIONE, A.; SZOLGAY, J.; KOHNOVÁ, S.; BLÖSCHL, G. Bayesian MCMC approach to regional flood frequency analyses involving extraordinary flood events at ungauged sites. *Journal of Hydrology*, v. 394, n. 1-2, p. 101-117, 2010. http://dx.doi.org/10.1016/j.jhydrol.2010.01.008.

GUSE, B.; HOFHERR, T. H.; MERZ, B. Introducing empirical and probabilistic regional envelope curves into a mixed bounded distribution function. *Hydrology and Earth System Sciences*, v. 14, n. 12, p. 2465-2478, 2010. http://dx.doi.org/10.5194/hess-14-2465-2010.

HASENACK, H.; WEBER, E. *Base cartográfica vetorial contínua do Rio Grande do Sul – escala 1:50.000*. Porto Alegre: UFGRS Centro de Ecologia, 2010. 1 DVD-ROM. (Série Geoprocessamento, 3).

HEO, J. H.; SHIN, H.; NAM, W.; OM, J.; JEONG, C. Approximation of modified Anderson-Darling test statistics for extreme value distributions with unknown shape parameter. *Journal of Hydrology*, v. 499, p. 41-49, 2013. http://dx.doi.org/10.1016/j.jhydrol.2013.06.008.

HOSKING, J. R. M.; WALLIS, J. R. Correction to " Some statistics useful in regional frequency. *Water Resources Research*, v. 31, n. 1, p. 251, 1995. http://dx.doi.org/10.1029/94WR02510.

HOSKING, J. R. M.; WALLIS, J. R. *Regional frequency analysis: an approach based on L-moments*. Cambridge: Cambridge University Press, 1997. 224 p.

INMET – INSTITUTO NACIONAL DE METEOROLOGIA. *Banco de Dados Meteorológicos para Ensino e Pesquisa: BDMEP*. Brasília, 2016. Disponível em: <http://www.inmet.gov.br/portal/index.php?r=bdmep/bdmep>. Acesso em: 14 set. 2016.

ISHAK, E. H.; RAHMAN, A.; WESTRA, S.; SHARMA, A.; KUCZERA, G. Evaluating the non-stationarity of australian annual maximum flood. *Journal of Hydrology*, v. 494, p. 134-145, 2013. http://dx.doi.org/10.1016/j.jhydrol.2013.04.021.

KHATIBI, R.; GHORBANI, M. A.; KASHANI, M. H.; KISI, O. Comparison of three artificial intelligence techniques for discharge routing. *Journal of Hydrology*, v. 403, n. 3-4, p. 201-212, 2011. http://dx.doi.org/10.1016/j.jhydrol.2011.03.007.

KJELDSEN, T. R.; SMITHERS, J. C.; SCHULZE, R. E. Regional flood frequency analysis in the Kwazulu-Natal province, South Africa, using the index-flood method. *Journal of Hydrology)*, v. 255, n. 1-4, p. 194-211, 2002. http://dx.doi.org/10.1016/S0022-1694(01)00520-0.

KUMAR, R.; CHATTERJEE, C. Regional flood frequency analysis using L-moments. *Journal of Hydrologic Engineering*, v. 10, n. February, p. 1-7, 2005. http://dx.doi.org/10.1061/(ASCE)1084-0699(2005)10:1(1).

KUMAR, R.; CHATTERJEE, C.; KUMAR, S.; LOHANI, A. K.; SINGH, R. D. Development of regional flood frequency relationships using L-moments for middle Ganga plains subzone 1(f) of India. *Water Resources Management*, v. 17, n. 4, p. 243-257, 2003. http://dx.doi.org/10.1023/A:1024770124523.

L'ECUYER, P.; SIMARD, R. TestU01: a C Library for empirical testing of random number generators. *ACM Transactions on Mathematical Software*, v. 33, n. 4, p. 22, es, 2007. http://dx.doi.org/10.1145/1268776.1268777.

MAKRIDAKIS, S.; WHEELWRIGHT, S.; HYNDMAN, R. J. *Forecasting methods and applications*. 3rd ed. New York: John Wiley & Sons, 1998.

MALEKINEZHAD, H.; NACHTNEBEL, H. P.; KLIK, A. Regionalization approach for extreme flood analysis using L-moments. *Journal of Agricultural Science and Technology*, v. 13, p. 1183-1196, 2011.

MERZ, B.; KREIBICH, H.; SCHWARZE, R.; THIEKEN, A. Review article "assessment of economic flood damage". *Natural Hazards and Earth System Sciences*, v. 10, n. 8, p. 1697-1724, 2010. http://dx.doi.org/10.5194/nhess-10-1697-2010.

NAGHETTINI, M. *Fundamental of statistical hydrology*. Switzerland: Springer International, 2017. 660 p. http://dx.doi.org/10.1007/978-3-319-43561-9.

NEYKOV, N. M.; NEYTCHEV, P. N.; VAN GELDER, P. H. A. J. M.; TODOROV, V. K. Robust detection of discordant sites in regional frequency analysis. *Water Resources Research*, v. 43, n. 6, p. n/a, 2007. http://dx.doi.org/10.1029/2006WR005322.

NOTO, L. V.; LA LOGGIA, G. Use of L-moments approach for regional flood frequency analysis in Sicily, Italy. *Water Resources Management*, v. 23, n. 11, p. 2207-2229, 2009. http://dx.doi.org/10.1007/s11269-008-9378-x.

RAO, R.; SRINIVAS, V. V. *Regionalization of watersheds: an approach based on cluster analysis*. Netherlands: Springer, 2008.

RIBEIRO, C. B. M.; MARQUES, F. A.; SILVA, D. D. Estimativa e regionalização de vazões mínimas de referência para a bacia do rio Doce. *Revista Engenharia Agrícola*, v. 13, p. 103-107, 2005.

SALAS, J. D. Analysis and modeling of hydrologic time series. In: MAIDMENT, D. R. (Ed.). *Handbook of hydrology*. New York: McGraw-Hill, 1992. cap. 19, p. 1-72.

SECKIN, N.; HAKTANIR, T.; YURTAL, R. Flood frequency analysis of Turkey using L-moments method. *Hydrological Processes*, v. 25, n. 22, p. 3499-3505, 2011. http://dx.doi.org/10.1002/hyp.8077.

SPAROVEK, G.; LIER, Q. D. J. V.; DOURADO NETO, D. Short Communication Computer assisted Koeppen climate classification: a case study for Brazil. *Royal Meteorological Society*, v. 27, p. 257-266, 2007.

TUCCI, C. E. M. *Regionalização de vazões*. Porto Alegre: Ed. Universidade/UFRGS, 2002. 256 p.

VEZZA, P.; COMOGLIO, C.; ROSSO, M.; VIGLIONE, A. Low flows regionalization in North-Western Italy. *Water Resources Management*, v. 24, n. 14, p. 4049-4074, 2010. http://dx.doi.org/10.1007/s11269-010-9647-3.

VILLARINI, G.; SERINALDI, F.; SMITH, J. A.; KRAJEWSKI, W. F. On the stationarity of annual flood peaks in the continental United States during the 20th century. *Water Resources Research*, v. 45, n. 8, p. 1-17, 2009. http://dx.doi.org/10.1029/2008WR007645.

VIVEKANANDAN, N.; SHUKLA, S. Flood frequency analysis using method of moments and L-moments of probability distributions. *Cogent Engineering*, v. 2, n. 1, p. 1-10, 2015. http://dx.doi.org/10.1080/23311916.2015.1018704.

WANG, Y.; LI, J.; FENG, P.; HU, R. A time-dependent drought index for non-stationarity precipitation series. *Water Resources Management*, v. 29, n. 15, p. 5631-5647, 2015. http://dx.doi.org/10.1007/s11269-015-1138-0.

ZHANG, Q.; GU, X.; SINGH, V. P.; XIAO, M.; XU, C. Stationarity of annual flood peaks during 1951-2010 in the Pearl River basin, China. *Journal of Hydrology* v. 519, p. 3263-3274, 2014. http://dx.doi.org/10.1016/j.jhydrol.2014.10.028.

ZRINJI, Z.; BURN, D. H. Flood frequency analysis for ungauged sites using a region of influence approach. *Journal of Hydrology*, v. 153, n. 1-4, p. 1-21, 1994. http://dx.doi.org/10.1016/0022-1694(94)90184-8.

Authors contributions

Felício Cassalho: Research, paper conception, data analysis, literature review, discussion of results, paper writing.

Samuel Beskow: Paper conception, advisor professor, discussion of results, paper writing.

Marcelle Martins Vargas: Data processing, data analysis, preparation of graphs and figures, discussion of results, manuscript revision.

Maíra Martim de Moura: Data processing, data analysis, preparation of graphs and figures, discussion of results, manuscript revision.

Leo Fernandes Ávila: Advisor researcher, discussion of results, manuscript revision.

Carlos Rogério de Mello: Advisor professor, discussion of results, paper writing.