Hydrological responses in equatorial watersheds indicated by Principal Components Analysis (PCA) – study case in Atrato River Basin (Colombia)

Respostas hidrológicas em bacias hidrográficas equatoriais indicadas pela análise de componentes principais (PCA) – estudo de caso na Bacia do Rio Atrato (Colômbia)

Sebastián Balbín Betancur[1], Didier Gastmans[1] (iD), Katherine Vásquez Vásquez[2], Lucas Vituri Santarosa[1] (iD),
Vinícius dos Santos[1] (iD) & Roberto Eduardo Kirchheim[3] (iD)

[1]Centro de Estudos Ambientais, Universidade Estadual Paulista "Júlio de Mesquita Filho", Rio Claro, SP, Brasil
[2]Laboratório de Hidrologia Florestal, Universidade de São Paulo, Piracicaba, SP, Brasil
[3]Serviço Geológico do Brasil, São Paulo, SP, Brasil

E-mails: sebastian.balbin@unesp.br (SBB), didier.gastmans@unesp.br (DG), kvasquezv@usp.br (KVV), lucas.santarosa@unesp.br (LVS), vinicius.santos16@unesp.br (VS), roberto.kirchheim@cprm.gov.br (REK)

ABSTRACT

The Atrato river basin is located in the Pacific fringe of Colombia, region with one of the highest precipitation rates in the world. The main purpose of this study is to determine the dominant processes in the hydrological responses along 17 sub-basins within the basin using principal component analysis. Watersheds located at the headwater presented a fast or medium response to the precipitation events, while higher flow homogeneity was observed in watersheds located at the lower portions of the basin. Three principal components were responsible for explaining 85.18% of the total variance. The component PC1 revealed the largest contributions for low flow behavior, being associated to precipitation, characteristic discharge values, compactness index, soil coverage and soil coarse textures. The component PC2 was assigned to the geological variables, fine and average texture soil and the average basin slope. Finally, the component PC3 has shown to be related to high flow patterns (maximum characteristic discharge values Q_5 and Q_1), igneous rocks and length of the basin. Highest specific discharge was associated to alluvial deposits and forest cover, whereas the slope was considered determinant for the run-off generation.

Keywords: Atrato river basin; Flow duration curves; Principal component analysis; Colombia.

RESUMO

A bacia do Rio Atrato está situada na costa Pacífica da Colômbia, região que apresenta uma das maiores taxas de precipitação no mundo. O principal objetivo desse estudo foi o de determinar os processos dominantes nas respostas hidrológicas de 17 sub-bacias situadas na bacia do Rio Atrato pela aplicação da análise das componentes principais (ACP). As sub-bacias localizadas nas cabeceiras do Rio Atrato apresentaram tempos de resposta médios/rápidos em relação aos eventos de precipitação, enquanto um fluxo bastante homogêneo foi observado nas bacias localizadas na porção baixa da Bacia. Três componentes principais explicam 85,18% da variância total. A componente principal CP1 está associado às contribuições para as baixas vazões, precipitação índice de compacidade e texturas finas e grossas do solo, A CP2 associado as variáveis fisiográficas relacionadas ao arcabouço geológico e as texturas finas e a CP 3 associado ao padrão das altas vazões Q_5 e Q_1 assim como a extensão da bacia. As maiores descargas específicas estão associadas a presença de depósitos aluviais e a cobertura florestal nas porções baixas da Bácia, enquanto as áreas montanhosas determinam a geração do fluxo superficial.

Palavras-chave: Rio Atrato; Curvas de permanência; Análise de componentes principais; Colômbia.

INTRODUCTION

In tropical and equatorial regions, the atmospheric convergence processes and moisture transport from the oceans are responsible for high rainfall events, which generate fast responses in water discharge in river channels (Strauch, 2013; Strauch et al., 2015). Climate change prediction models (Gleeson et al., 2020) run for equatorial and tropical regions revealed a decrease in rainfall rates. Catchments located in these regions are, therefore, highly susceptible. Changes in their hydrological regime may impact the provision of regulatory ecosystem services.

Undoubtedly, it is a fact of relevance, to be taken into consideration within regional and local water resource management policies (Meixner et al., 2016; Ng et al., 2010; Sophocleous, 2002).

The Colombian territory is affected by three major sources of humidity: the Caribbean Sea, the Pacific Ocean and the Amazon rainforest. It presents a complex climate configuration, with remarkable spatiotemporal variations of precipitation associated to a large physiographic variability. Under such circumstances, the prediction and estimation tasks for hydroclimatological variables become complex (Poveda Jaramillo, 2004). In addition, there are other local factors such as the orography (with a maximum altitude of 3,945 masl) and climate features, such as El Niño Southern Oscillation (ENSO) among others, that may interfere on rainfall intensity and seasonality (Rueda & Poveda Jaramillo, 2006; Garreaud, 2009; Durán-Quesada et al., 2012). Recent simulations revealed water scarcity risk scenarios for the Andean region and flooding in the coastal regions associated with sea level rise (Intergovernmental Panel on Climate Change, 2014a, 2014b; Hurtado & Mesa, 2015; Pardo & Alfonso, 2018).

The Colombian Pacific Region presents the highest precipitation rates observed in the South America. They are associated with some climatic features, such as the migration of the Intertropical Convergence Zone (ITCZ), the Low Level Jet Stream of the Colombian West Region (Chorro del Occidente Colombiano, CHOCO), the meso-scale convective systems and an the orographic barriers represented by the Andes (Poveda Jaramillo & Mesa, 2000; Poveda Jaramillo et al., 2014). The Atrato River Basin, located in the northwestern region of Colombia, is mostly influenced by the moisture originated in the Pacific Ocean and the Caribbean Sea (Instituto de Investigaciones Ambientales del Pacífico, 2013; Poveda Jaramillo et al., 2014). Due to these high precipitation rates, the Atrato River presents the highest specific discharge among all the other rivers in Colombia (Leyva, 1993).

Actually, besides bearing heavy rainfall and high seismicity, the Atrato River Basin (hereafter ARB) is facing increasing demands for natural resources (fauna and flora) and several conflicts related to land use, illegal mining and uncontrolled deforestation (Vélez & Aguirre, 2016). These stresses increase the socioeconomic vulnerability of the population living in the basin and the occurrence of natural disasters, such as flooding and landslides (Mosquera Machado, 2006).

Due to its complexity, the ARB is considered an ideal target for multidisciplinary scientific research to assess hydroclimatological dynamics (Garreaud, 2009; Célleri & Feyen, 2009) and to provide knowledge to cope with climate change impacts.

The hypothesis here is that in such a complex environment, the hydrologic properties of the basin, such as water storage, water movement and baseflow characteristics are highly affected by the watershed topography and geomorphology (Price, 2011). Additionally, according to (Cheng et al., 2012; Coopersmith et al., 2012), the landscape framework is shaped directly by the regional climatic conditions, which, also determines the response of stream flows to rainfall intensity (Berhanu et al., 2015).

Among the available hydrological tools to evaluate the response of stream flows to the physical and climatic variables, the evaluation of Flow Duration Curves (FDC) is a widely accepted methodology with recognized practical efficiency (Ilstedt et al., 2016). It is based on graphical and statistical methods to identify the influence of the physical and climatic variables of the basin and how far they describe its hydrological behavior (Smakhtin, 2001; Brown et al., 2005). Water flow impact assessments using the FDC approach, considering forest cover, soil and geology framework, are offered by Best et al. (2003); Brogna et al. (2017); Peña-Arancibia et al. (2019).

Due to the large number of variables that affect the FDC responses, multivariate techniques can be applied precisely to identify the most dominant ones in hydrological processes. In these cases, it is recommended to use cluster analysis and/or the clusters obtained from the Principal Component Analysis (PCA) (Galbraith et al., 2002; Abdi & Williams, 2010).

The application of PCA techniques has the great advantage since it reduces the number of the variables used to explain the hydrological behavior, by retaining only those principal components that explain the most significant portion of the data variance (Westra et al., 2007). It is a way of complementing the information obtained through the FDC analysis.

The PCA techniques have been used in hydrology: (i) to identify cause-effect relationships; (ii) to detect redundant variables; (iii) to further identify zones with homogeneous climatic or hydrological characteristics, and also; (iv) to perform analysis of morphometric variables in watersheds (Richman, 1981; Westra et al., 2007; Singh et al., 2009; Sharma et al., 2015).

These techniques have been applied in several watersheds located in the mountainous Andes. Crespo et al. (2011) and Masiokas et al. (2019) were able to identify a differentiation between basins based on hydrological series and types of soils and also patterns of spatial variability on runoff using monthly river discharge values respectively. The study of Boscarello et al. (2016) discussed the morphometric and climatological key drivers on hydrological dynamics.

Despite the hydrological importance of the watersheds located in Andean regions, such as the Atrato river basin, few studies had discussed the impacts of land use changes and how the hydrological knowledge could be improved by incorporating them in the overall analysis (Ponette-González et al., 2014).

In order to understand the hydrological controls of the Atrato river flow a key scientific question has been proposed: What are the most influential parameters that control the hydrological response of the sub-basins of the Atrato river and how to determine them? Having this question as a guideline, the objective of the present study is to identify and to characterize the physiographic and morphological variables that control the hydrological responses of the Atrato river sub-basins, based on data from a long term monitoring program (1984-2017).

This study is pioneering as it represents the first attempt to assess the relationships between hydrological and physiographic variables using statistical multivariate approach for the ARB. The results will provide an understanding of land use and vegetation cover on water availability in the region, therefore, pillars for the preservation and conservation of the ecosystems and for the integrated water management policies.

STUDY AREA

The ARB is located in the northwestern portion of the Colombian territory (Figure 1A), in the Chocó biogeographic region. It occupies an area of about 35,000 km² from its headwaters, located in the western Colombian Andes, to its mouth in the Caribbean Sea (Figure 1B).

The geological framework of the basin consists of a sedimentary sequence represented mainly by marine deposits, which can reach thicknesses of 10,000 m, deposited between the Lower Eocene and Pliocene, hereafter sediments, over a Cretaceous igneous-sedimentary basement, hereafter denominated basement. This sequence is covered by silt-clay and sandy alluvial deposits with high organic matter content, without cementation and poorly consolidated, hereafter alluvial (Alcárcel & Gómez, 2017).

This complex geological framework can be simplified in 3 subgroups according to the lithological and hydrogeological characteristics of the outcropping sequences (Figure 1D): A unit where sedimentary rocks prevail (48%), followed by a unit composed of Quaternary deposits (32%), and finally, a unit built by igneous formations (20%).

The soils of the ARB are of mainly medium-fine, medium and coarse-medium textures (Figure 1E). Forests and natural cover represent 80% of the ARB area (Figure 1C), followed by agricultural areas with 15.6%. It is noteworthy that only 0.3% of the basin can be considered anthropized, essentially due to urban settlements, mining projects, roads and industrial districts (Instituto Geográfico Agustín Codazzi, 2007).

The discharge at the mouth of the Atrato River is approximately 4,900 m³.s⁻¹, with a specific discharge of 0.141 m³.km⁻².s⁻¹, that is directly related to the average rainfall of 4,600 mm.year-1 with an average temperature of 24.3°C.

The high precipitation rates observed are a result of the orographic effects over the moisture coming from the Pacific Ocean (Guzmán et al., 2014; Poveda Jaramillo et al., 2014), which are then generated due to permanent convective activity in the region throughout the year (Zea et al., 2000), as well as the influence of the "CHOCO jet" presents larger magnitudes (Hoyos et al., 2018). This climatic and orographic effect combination is responsible
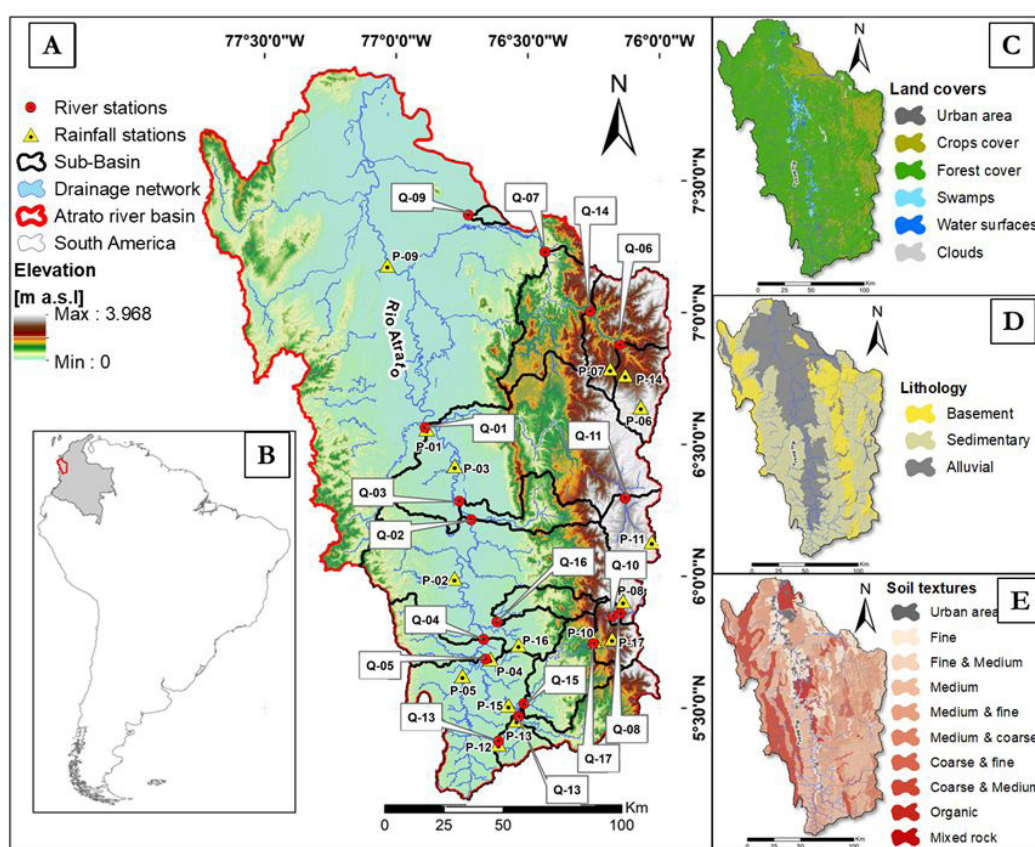

Figure 1. Location and physiographic variables of the Atrato river basin. A) Relief (modified from Japan Aerospace Exploration Agency, 2017), with the location of river discharge gauging station, indicated by letter Q and number; B) Location of the basin in the South America and Colombian territory; C) Map of land cover (modified from Instituto de Hidrología, Meteorología y Estudios Ambientales, 2014); D) Simplified geological map. Lithology are grouped in the three main groups (modified from Alcárcel & Gómez, 2017), and E) Soil textures (modified from Instituto Geográfico Agustín Codazzi, 2007).

for the existence of different types of hydrometeorological cycles, in regard to precipitation and to river discharge as well (Urrea et al., 2019).

MATERIAL AND METHODS

The ARB was divided in 17 sub-basins (Figure 1), which had been monitored by gauging stations. Discharge records on a daily time scale are practically continuous, with periods ranging from 17 to 46 years of observation and with a percentage of missing data between 0% and 20%. Morphometric parameters and physiographic variables are presented in Table 1. In addition to the fluviometric stations, rainfall data from 17 stations, one for each sub-basin, with daily readings and records ranging from 22 to 71 years and 0 to 16% missing rate, were considered in the analyzes

For the sake of data homogenization, a common period of hydrometeorological observations between 1984 and 2017 was selected. The fluviometric and rainfall data were obtained from the Water Resource Information System (SIRH) of the Colombian Institute of Hydrology, Meteorology and Environment (IDEAM) (Instituto de Hidrología, Meteorología y Estudios Ambientales, 2014).

The most important morphometric, physiographic and discharge variables that seemed to build up hydrological responses at the ARB were evaluated. The morphometric variables considered here were: area (A), perimeter (Pe), compactness index or Gravelius index (K) (Eagleson, 1970), basin length (LB) and average basin slope (ABS). A and P values were discarded from subsequent multivariate statistical analyzes due to their redundancy with K. Physiographic variables were defined from the land cover mapping and geological map (Table 1 and Figure 1).

Table 1. Morphometric parameters of the study basins.

Id	Sub-Basin	Area (A) [km²]	Perimeter (Pe) [km]	K index	Length of the basin (L.B) [km]	Average slope of the basin (A.B.S) [%]	Lithology [% area] Basement (B)	Sediments (S)	Alluvial (Al)	Cover land [% area] Crops (C)	Forest (F)	Others (O.C)	Soil texture [% area] Fine (F.T)	Average (A.T)	Coarse (C.T)	Other (O.T)	Annual precipitation (P) [mm/year]
1	Bellavista	15,560	703.0	1.6	157.1	20.0	15.9	63.9	20.3	14.0	82.0	4.0	9.3	75.4	12.6	2.7	5416.7
2	Tagachi	9,865	568.0	1.6	117.6	18.2	13.3	66.4	20.3	10.5	85.0	4.6	5.7	78.9	13.3	2.1	6254.0
3	San Antonio Padua	10,504	621.0	1.7	125.3	18.2	15.0	64.3	20.7	10.7	84.8	4.5	6.2	78.8	12.7	2.2	6149.8
4	Belen	5,426	517.0	2.0	78.1	18.8	14.5	73.2	12.3	14.1	82.1	3.8	3.9	83.0	11.6	1.5	6395.1
5	Quibdo-Automatica	4,777	462.0	1.9	74.8	19.4	16.5	73.2	10.3	13.3	83.1	3.7	3.2	86.7	8.6	1.4	6364.2
6	El Añil	714	128.0	1.4	43.3	40.2	16.6	71.6	11.8	64.6	34.7	0.7	35.5	63.9	0.0	0.7	2193.0
7	Mutata	3,331	345.0	1.7	90.8	40.1	41.9	54.0	4.1	46.7	51.8	1.5	20.3	78.1	1.1	0.5	2863.9
8	El Siete	208	66.0	1.3	17.8	40.9	15.4	84.6	0.0	37.1	62.2	0.7	0.0	99.9	0.0	0.1	3156.0
9	Bajira	78	43.0	1.4	17.2	1.6	0.0	0.0	100.0	88.4	10.8	0.8	37.7	62.1	0.0	0.2	4190.0
10	Pte Las Sanchez	226	72.0	1.4	20.0	41.4	14.2	85.8	0.0	38.4	60.8	0.8	0.0	99.9	0.0	0.1	3230.0
11	La Magdalena	631	129.0	1.5	33.1	35.2	28.8	68.7	2.5	28.4	70.3	1.3	0.0	97.1	2.9	0.0	2947.1
12	Aguasal	911	205.0	1.9	61.6	32.5	35.8	64.2	0.0	13.5	82.8	3.7	1.8	93.8	3.5	0.9	8404.7
13	Pte Certegui	262	77.0	1.3	23.2	6.7	0.0	100.0	0.0	0.3	96.3	3.4	1.0	98.0	0.0	1.0	6985.2
14	Dabeiba 2	1,968	221.0	1.4	60.6	44.8	21.8	73.9	4.3	62.3	37.3	0.4	18.2	81.4	0.0	0.4	2286.5
15	Gindrama	1,637	231.0	1.6	71.7	29.7	27.9	67.6	4.5	21.6	76.9	1.5	2.5	82.9	14.1	0.5	6259.4
16	Negua	422	135.0	1.9	47.1	21.9	30.3	64.2	5.5	5.7	92.7	1.7	4.9	90.0	4.6	0.6	6650.2
17	Arrayanes	9,865	568.0	1.4	117.6	18.2	28.8	68.7	2.5	0.3	96.3	3.4	0.0	97.1	2.9	0.0	4129.5

Morphometric variables *Physiographic variables*

The FDCs (Flow Duration Curves) were constructed with the flow series expressed as water column on daily time scale (mm. day^{-1}), aiming to remove the effect of the basin area and allowing the comparison between the curves, as well as the evaluation of the hydrological responses based on the characteristic discharge values of the basins, (Q_1, Q_5, Q_{10}, Q_{50}, Q_{70}, Q_{90} and Q_{95}), where the number subscript correspond to the percentage of time where the discharge is exceeded. . The Ratios Q_{90}/Q_{10} and Q_{90}/Q_{50} were also calculated since they are considered indicators of maximum and minimum flows, respectively (Smakhtin, 2001).

Statistical analysis

A trend analysis was performed in order to identify changes in the mean, median of the data series (precipitation and discharge), as well as to verify the existence of trends or randomness that could affect the subsequent data analysis (Castro & Carvajal, 2010; Guajardo & Granados, 2010). The Trend Analysis 1.2 Software, developed by Cooperative Research Centre for Catchment Hydrology (CRCCH) of Australia (Chiew & Siriwardena, 2005) has been used for this purpose. Analysis included nonparametric statistical tests for all stations data series: (i) The Mann-Kendall test to assess the existence of trends; (ii) The CUSUM Free Distribution test to evaluate the average jumps; (iii) The Rank-Sum test to evaluate the differences in the median, where levels of significance allowed the identification of the null hypothesis acceptance evidences (H_0), defined as no trend, average jump, or median difference, respectively.

Correlation between the characteristic discharge values, morphometric and physiographic variables were calculated in order to verify the dependency of these variables, and if some of the correlation observed could be associated to the hydrological response of the watersheds. Initially, the Kolgomorov-Smirnov test was applied to evaluate whether the data presented a parametric or nonparametric distribution. As the result indicated a parametric distribution for the set of variables, the Pearson correlation test was then used, and the correlations were calculated using the R Heatmaply package (R Core Team, 2018). Heatmaply also deploys a seriation package to build a dendogram structure using a default option OLO (optimal leaf ordering) for optimization purposes.

Due to the great number of variables used in this study, the principal component analysis (PCA) has been applied to reduce the large initial set of correlated (or not) variables into a new orthogonal and uncorrelated set of factors called principal components (PC), that account to explain the variance in the original data set. This new representation has compressed the data and kept only the most important information, turning the data systematization into a more simplified exploratory task (Legendre & Legendre, 2003; Jackson, 1991; Abdi & Williams, 2010).

Original variables were initially normalized in order to remove the effect of measurement units, rendering the data dimensionless, as expressed below:

$$Z_i = (X_I - \mu)/\sigma \qquad (1)$$

Zi is the standardized variable i, Xi is the original variable i, μ is the mean and σ is the standard deviation of the dataset (Davis, 1973).

As soon as the data has been normalized, the PCA analysis involved the calculation of the correlation matrix R and the determination of the eigenvalues ($\lambda 1$, $\lambda 2$,…, λp, where p is the number of original variables), and the corresponding eigenvectors (a1, a2, …, ap), through the matrix equation, where "I" is the identity matrix:

$$|R - I\lambda| = 0 \qquad (2)$$

The composition of the principal components was performed by the application of matrices of eigenvectors as the factors in a linear combination of standardized variables (Noori et al., 2010).

These principal components are orthogonal to each other and are arranged in a descending order based on the amount of variance explained. Thus, the first principal component explains the higher amount of the total variance from the original variables, while the second one captures the higher proportion of the total variance, which is not encapsulated by the first one and so, further on. The retention of meaningful components is done on the basis of the sum of total explained variance.

Eigenvalues and eigenvectors were extracted from the covariance matrix of the original variables using the STATISTICA® software. The components that explained the variance of the data set were identified and ranked hierarchically into groups with the same dominant component. The correlations obtained from the PCA analysis were compared and the variables holding the greatest influence on the hydrological behavior of the different basins were finally defined.

RESULTS AND DISCUSSION

Trend and statistical analysis

No significant trends were identified for the precipitation series in 71.4% of the evaluated rain gauging stations. In the remaining stations, the statistical analysis indicated the existence of two trends in the average precipitation magnitudes: (i) An increasing trend comprised of three stations; (ii) A decrease trend in the other two stations showing different significance levels (with 3.1% of cases rejecting H_0 with a significance for $\alpha = 0.10$, for $\alpha = 0.05$ in 8.3% of cases and for $\alpha = 0.01$ in 17.2% of cases). Statistical test of changes in averages indicated that in 43% of the rain gauges stations no changes have been observed since and passed the H_0 of the CUSUM free distribution test. On the other hand, in 23% of the stations the H_0 test could be rejected with a significance level of $\alpha = 0.10$, indicating slightly evidences to accept or reject H_0; 14% and 20% of stations, respectively, rejected H_0 with $\alpha = 0.05$ and $\alpha = 0.01$, indicating that these stations had a nonparametric behavior, corroborated by the existence of trends.

In 59% of the river gauging series there was no observed trend ; the remaining stations (41%) showed positive trends (indicated by an increase in the average flow magnitudes); 12% revealed statistical significance with little evidence of trends ($\alpha = 0.10$); 17% had average evidence of trend ($\alpha = 0.05$); and 12% showed strong indications for tendency ($\alpha = 0.01$) in the series, specifically the El Siete and Negua stations (Table 2).

Table 2. Homogeneity tests for series of precipitation and river discharge gauging stations.

Sub-basin	Mann-Kendall		CUSUM Free Distribution		Rank-Sum	
	Q	**P**	**Q**	**P**	**Q**	**P**
1	S (0.1)	NS	S (0.1)	S (0.1)	NS	S (0.05)
2	S (0.1)	NS	S (0.1`)	S (0.1)	NS	NS
3	S (0.05)	NS	NS	NS	NS	NS
4	NS	NS	NS	NS	NS	NS
5	NS	NS	NS	NS	NS	S (0.05)
6	NS	NS	NS	NS	S (0.1)	NS
7	NS	NS	NS	NS	NS	NS
8	S (0.01)	S (0.01)	S (0.05)	S (0.01)	S (0.05)	S (0.01)
9	NS	NS	NS	S (0.1)	NS	NS
10	S (0.05)	NS	S (0.1)	NS	NS	NS
11	NS	NS	NS	NS	S (0.1)	NS
12	NS	NS	S (0.1)	NS	S (0.01)	S (0.05)
13	NS	S (0.01)	NS	S (0.01)	NS	S (0.01)
14	S (0.05)	NS	S (0.05)	NS	S (0.05)	NS
15	NS	NS	NS	S (0.1)	NS	NS
16	S (0.01)	NS	S (0.05)	NS	NS	NS
17	NS	S (0.05)	NS	S (0.1)	NS	NS

NS: non-significance to reject H_0; S: significance to reject H_0, with $\alpha = 0.1$: little evidence to reject H_0; $\alpha = 0.05$: possible evidence to reject H_0, $\alpha = 0.01$: strong evidence to reject H_0; Q: Discharge station; P: Precipitation station.

It was also shown that 59% of the discharge series did not present evidences of changes; while 24% of the stations revealed slight evidence of average changes for significance levels $\alpha = 0.10$; and 18% presented changes for significance levels for $\alpha = 0.05$ (Table 2).

Rain and river gauging stations with trends of changes are situated in the same region, where nonparametric precipitation behaviors were detected. Nevertheless, there are also some river stations where discharges had increased regardless of precipitation patterns. This behavior cannot be directly attributed to climate change and there are errors associated to the monitoring program and analysis.

Characteristic discharge values for the study sub-basins and hydrological indicators were estimated from the FDC's (Table 3). Large variation in the magnitude of the characteristic discharges were observed. Variations reached more than 10 times to Q1 and about 5-6 times for the other characteristic discharges. Despite the similarity observed in slopes and trends for most of the FDC's curves, with exception of sub-basins 9 and 13 (single group with a high slope in discharge values, blue line in Figure 2), the other basins can be grouped in two groups based on their characteristic discharge values. The first group, representing the headwater basins, presented lower values for the characteristic discharge values (brown lines in Figure 2), indicating a low water release associated to the precipitation events. The second group, located at the lower part of the river, is characterized by higher characteristic discharges values (green lines in Figure 2); their flow rates are higher and more stable over time, which could be related to slow responses to the precipitation events and a constant pattern of aquifer contribution.

The hydrological and physiographic variables can be divided in two groups, according the dendogram presented in Figure 3. The first group (green part of the dendogram) encompasses the following variables (Al) alluvium coverages, (S) sedimentary
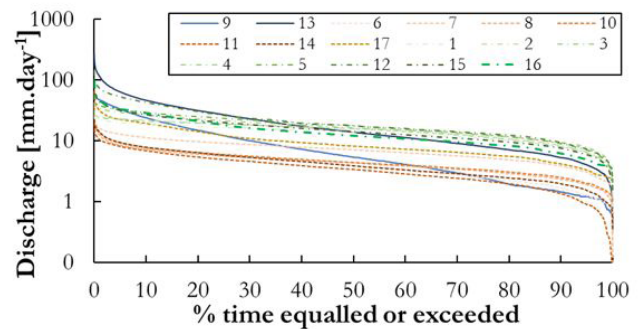

Figure 2. Flow Duration Curves (FDC) Group I represented in brown; Group II in green, and Group III in blue. Location of gauging stations are presented in Figure 1, and details about groups division on the main text.

rocks; (B) basement, average and fine texture soil cover (AT and FT), crops (C) and average basin slope (ABS). The second group of variables (red part of the dendogram) encompasses the morphometric variables (K index and length of the basin (LB), mean annual precipitation (P), coarse and other soil textures (CT and OT), forest and other soil coverage (F and OC) and the characteristic discharge values, which were assembled in two sub-groups: (i) lower discharge values (Q50, Q70, Q90 and Q95), associated to other coverages (OC) and K index; (ii) higher discharge values (Q10, Q5 and Q1).

The Pearson correlation between precipitation and characteristic discharge values can be assumed to be valid, because no significant trends in precipitation and discharge series were observed, and the values indicated a positive correlation, ranging from 0.70 to 0.93 between average precipitation (P) and characteristics discharge (Q95-1). A higher correlation was

Table 3. Characteristics discharge values presented in mm.day^{-1} from FDC's curves.

| Sub-basin	Q_{95}	Q_{90}	Q_{70}	Q_{50}	Q_{10}	Q_5	Q_1	Q_{90}/Q_{10}	Q_{90}/Q_{50}
1	6.27	8.06	11.74	14.08	18.78	20.15	22.38	0.43	0.57
2	6.51	8.45	12.72	15.81	21.86	23.22	25.19	0.39	0.53
3	7.82	9.42	13.40	16.06	21.33	22.54	24.14	0.44	0.59
4	6.42	8.06	12.58	16.29	26.56	29.62	34.30	0.30	0.49
5	7.07	9.04	13.82	17.65	29.08	32.25	38.47	0.31	0.51
6	1.27	1.69	2.82	3.95	7.44	8.62	13.19	0.23	0.43
7	3.06	3.86	5.68	7.14	11.33	12.93	16.02	0.34	0.54
8	1.91	2.33	3.43	4.35	7.30	8.60	12.55	0.32	0.54
9	1.22	1.44	2.94	5.38	24.20	33.73	47.46	0.06	0.27
10	1.99	2.48	3.54	4.43	7.84	9.36	13.85	0.32	0.56
11	0.90	1.36	2.43	3.40	6.91	8.12	10.60	0.20	0.40
12	7.34	9.44	14.19	17.86	43.39	53.53	73.98	0.22	0.53
13	4.05	5.29	9.10	13.95	46.60	64.21	117.07	0.11	0.38
14	1.53	1.97	2.89	3.90	7.84	9.43	14.00	0.25	0.51
15	5.95	7.60	11.09	14.29	26.98	31.86	44.32	0.28	0.53
16	4.91	6.22	9.64	12.18	28.77	33.62	47.19	0.22	0.51
17	3.31	4.20	6.50	8.21	19.35	22.66	31.81	0.22	0.51
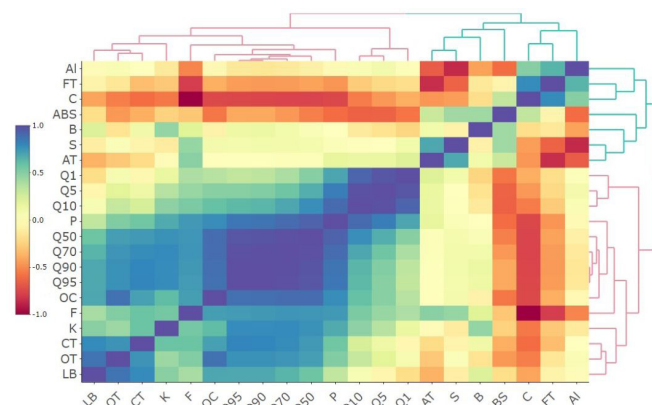
Figure 3. Correlation matrix and dendrogram. Lithology variables: (Al) Alluvial, (S) Sediments; (B) Basement. Soil Textures variables: (F.T) Fine; (A.T.) Average; (C.T.) Coarse; (O.T) Other. Soil Coverage: (C) Crops; (F) Forest; (O.C) Other. Basin shape variables: (ABS): Average Basin Slope; (LB) Basin Length; (K) Compactness index. Mean annual precipitation (P), Characteristic discharge values: (Q95); (Q90); (Q70); (Q50); (Q10); (Q5) and (Q1).

observed with Q50 (r=0.93), with tendency to decrease towards greater and lower discharges.

Some correlations between characteristic discharge values and physiographic variables were also identified (Figure 3). The Q95-50 presented high correlation with the K index and LB, respectively about 0.78 and 0.65. On the other hand, no correlation was observed with Q1-10. Regarding ABS a negative correlation was observed, increasing to higher characteristics discharge.

Land use patterns and soil textures seem to represent an important control factor on river discharges: (i) Forest and other soil coverage areas resulted in a positive correlation, decreasing towards higher discharge (0.75<r<0.39 and 0.88<r<0.36, respectively); (ii) Inverse correlation with crop areas, increasing towards higher characteristic discharge values (-0.39<r<-0.75): (iii) Areas covered by fine textures soils do not show strong correlation with the characteristic discharge values, increasing towards higher characteristic discharge (-0.20<r<-0.45); (iv) The soil coarse textures rendered a strong correlation with lower characteristic discharge values, indicating the storage role played by this soil and the discharges during low rainfall periods; (v) No important correlation with geological framework has been observed (Figure 3).

Principal Components Analysis (PCA)

The PCA showed that the first three components were responsible by explaining 85.18% of the total variance of the data (Table 4). The principal component 1 (PC1), which accounted for 51.35% of total variance, had the largest influence on the characteristic discharge values (P, Q_{95}, Q_{90}, Q_{70}, Q_{50} and Q_{10}), followed by the K index, associated with forest (F),other soil textures (OT) and soil coarse textures (CT). These variables were grouped in the red portion of the dendrogram (Figure 3).

The principal component 2 (PC2) explained 18,84% of the total variance. The variables presenting higher loads were those related to the geological features (areas covered by sedimentary rocks and alluvial deposits), fine and average texture soil and the average basin slope (ABS), (green portion of the dendrogram at the Figure 3). Finally, the principal component 3 (PC3) explained 14.99% of the total variance and showed the highest weights associated with the maximum characteristic discharge values (Q_5 and Q_1), basement (B) and length of the basin (LB), (red portion of the dendrogram except for basement at the Figure 3).

The variables associated with PC1 were considered the most determinant for the flow control in the sub-basins presenting maximum specific discharge along the historical series. Discharge, in this case, is due to land cover occupied by forested areas and coarse soil textures. This factor is associated to the sub-basins belonging to the Group II of FDCs (green lines at Figure 2 and green triangles at Figure 4A). The PC1 showed a positive load of eigenvalues for the magnitudes of the specific discharges and the precipitation (Table 4).

Table 4. Maximum variance explained by the three main factors. Values in bold indicate the higher scores variables for the PC (For explanation of variables refer to Figure 3).

Type	Variable	PC1	PC2	PC3
Morphometric	K	**0.74**	0.05	0.24
	LB	0.59	0.26	**0.69**
	ABS	-0.53	**-0.61**	0.40
Hydrographic	P	**0.93**	0.06	-0.29
	Q_{95}	**0.96**	0.07	0.19
	Q_{90}	**0.96**	0.05	0.18
	Q_{70}	**0.98**	0.07	0.11
	Q_{50}	**0.98**	0.10	-0.01
	Q_{10}	**0.75**	0.12	-0.61
	Q_5	0.65	0.12	**-0.72**
	Q_1	0.50	0.03	**-0.82**
Physiographic	Crops (C)	**-0.86**	0.40	-0.03
	Forest (F)	**0.85**	-0.43	0.02
	Other cover (OC)	**0.91**	0.14	0.11
	Fine texture (FT)	-0.54	**0.74**	0.03
	Averange texture (AT)	0.13	**-0.88**	-0.30
	Coarse texture (CT)	**0.74**	0.19	0.49
	Other texture (OT)	**0.73**	0.30	0.38
	Basement (B)	0.07	-0.31	**0.44**
	Sediments (S)	0.15	**-0.84**	-0.04
	Alluvial (Al)	-0.17	**0.91**	-0.19
	Eigen value	10.78	3.96	3.15
	% explained variance	51.35	18.84	14.99
	% cumulative variance	51.35	70.19	85.18

PC1 – Principal Component 1; PC2 – Principal Component 2; PC3 – Principal Component 3.

Figure 4. Distribution of the basins according to the Principal Components (A) PC 1 (average and minimum flow dominant) versus PC 2 (geological aspects, fine and average texture), (B) PC 2 (geological aspects) versus PC 3 (maximum flow dominants), (C) PC 1 (average and minimum flow dominant) versus PC 3 (maximum flow dominants).

On other hand, the PC2 was interpreted as responding for the discharge processes and has hosted the variables related to the geological settings (Figure 4B). Most of the sub-basin belonging to Group II are characterized by a higher percentage of alluvial deposits, except sub-basins 12 and 16 which presented remarkable areas of basement rocks, with a hydrological response similar to observed in the sub-basins belonging to the Group I. As a way of verifying the relationship between geological variables and minimal discharges, some authors have suggest detailing the geological discretization of the study area (Carlier et al., 2018).

Sub-basins 14, 6 and 7, despite the fact of being located in the upper part of the ARB and presenting an opposite hydrological response, seemed to have flows controlled by the geological framework. The predominance of alluvial areas in these sub-basins could attenuate the hydrological responses and make these sub-basins to behave similar to the ones of group II. Finally, similar to PC1, the PC3 was associated with sub-basins in which the runoff is dominant (Figure 4C).

The flow magnitudes presented in the permanence curves (Figure 2) may also be directly related to the principal components. The sub-basins of the Group I showed negative PC1 and PC3 values, indicating lower values for the specific discharge. They are constituted by higher percentage of agricultural land and sedimentary deposits with higher ABS. Conversely, the Group II sub-basins presented
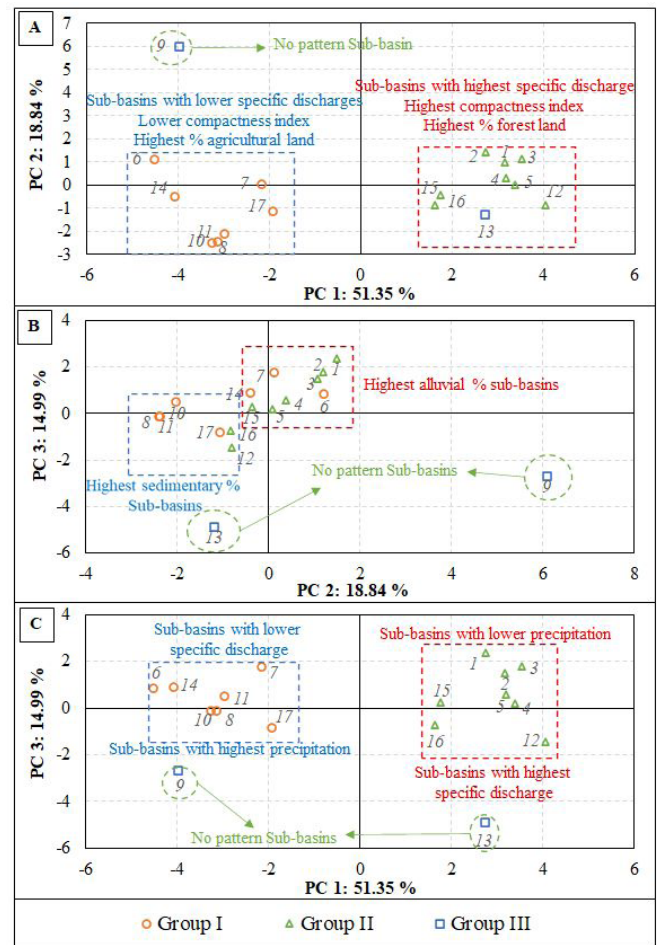
positive PC1 and PC3 values, indicating higher specific discharge. They present higher percentage of forest areas and alluvial deposits and do have lower ABS and greater magnitude in precipitation, compared to the previous ones. The PC1 maximum loads of variables were associated to the characteristic discharge values and to the forest cover. On other hand, the areas dominated by agricultural land cover seemed to favor the runoff generation. According to Le Maitre et al. (1999), agricultural areas promote changes in the macro and microstructure of soils preventing water infiltration and other processes.

The relationships between fine, average textures and the ABS, responsible for building up higher negative loads for the PC2, lead to slow discharges, as previously stated by Brown et al. (2013). These authors observed small changes in lower discharges in watershed dominated by fine soils and lower slopes in contrast to watersheds dominated by coarser soils and higher slopes. On the other hand, areas bearing high hydraulic conductivities coincident with areas presenting lower to average slopes seemed to facilitate

water infiltration and storage, increasing the permanence of minimum flows and base flow (Smakhtin, 2001).

In PC3, the amount of surface runoff is evidenced as a determinant in the composition of maximum flows in areas with predominance of basement rocks. In comparison with areas dominated by other lithologies, related to PC2, the basement rocks tend to decrease the soil water infiltration capacity, facilitating the surface runoff and delivering higher magnitudes of maximum flow rates. The sub-basin specific discharge can be analyzed through the different types of FDC's (Figure 2). The sub-basins with type II curves that are predominantly covered with forest seemed to present a high production of water leading to a more regulated discharge pattern. This is due to the control exerted by interception and evapotranspiration, which tends to homogenize the rain arrival in the soil and subsequent infiltration and water release. The sub-basins, which are characterized by the type I FDC's, situated in higher slopes areas, are covered by agricultural activities and do have the lowest specific discharge. In agricultural areas, due to the lack of vegetation the river flow responses to the precipitation events are faster and therefore its specific performance is lower (Brown et al., 2013).

CONCLUSIONS

In the studied area, the specific discharge variations are mainly controlled by land cover (forest and agriculture), the presence of coarse soils, and, to a lesser extent, by the geological framework (alluvial deposits versus sedimentary rocks). Areas covered by forest associated to coarse soil presented the higher capacity to regulate low flow. On the other hand, sub-basins with intense agricultural activities presented higher instantaneous discharge, due to higher runoff as a combined effect of evapotranspiration rate reduction and the roughness of the terrain.

Despite the low correlation, the geological setting is an important discharge control factor, as evidenced by the grouping of variables from PC2. However, to corroborate this relationship, a more detailed geological assessment is required. The geological framework of the ARB also plays an important role on discharge. The sub-basins situated on more permeable rocks showed higher specific discharge. This is an important aspect to be considered when it comes to water management practices and policies under climate changes scenarios.

Despite the concordant results between FDCs and PCA analysis, the FDCs analysis must be understood as a preliminary assessment of the key factors related to the hydrological responses. Complex areas, such as the ARB, require caution in the application of FDCs, since some simplifications could affect subsequent analyzes.

The statistical support given by the PCA analysis has minimized biases in the identification of such factors. The methodology used within this study proved to be a robust tool for understanding the hydrological response and water performance of complex and heterogeneous regions.

REFERENCES

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(4), 433-459. http://dx.doi.org/10.1002/wics.101.

Alcárcel, F. A., & Gómez, J. (2017). *Mapa geológico de Colombia escala 1:2.000.000*. Bogotá: Servicio Geologico Colombiano.

Berhanu, B., Seleshi, Y., Demisse, S., & Melesse, A. (2015). Flow regime classification and hydrological characterization: a case study of Ethiopian rivers. *Journal Water*, *7*(12), 3149-3165. http://dx.doi.org/10.3390/w7063149.

Best, A. E., Zhang, L., McMahon, T. A., & Western, A. W. (2003). *Development of a model for predicting the changes in flow duration curves due to altered land use conditions* (pp. 861-866). Canberra: Cooperative Research Centre for Catchment Hydrology.

Boscarello, L., Ravazzani, G., Cislaghi, A., & Mancini, M. (2016). Regionalization of flow-duration curves through catchment classification with streamflow signatures and physiographic-climate indices. *Journal of Hydrologic Engineering*, *21*(3), 05015027. http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0001307.

Brogna, D., Vincke, C., Brostaux, Y., Soyeurt, H., Dufrêne, M., & Dendoncker, N. (2017). How does forest cover impact water flows and ecosystem services? Insights from "real-life" catchments in Wallonia (Belgium). *Ecological Indicators*, *72*, 675-685. http://dx.doi.org/10.1016/j.ecolind.2016.08.011.

Brown, A. E., Western, A. W., McMahon, T. A., & Zhang, L. (2013). Impact of forest cover changes on annual streamflow and flow duration curves. *Journal of Hydrology*, *483*, 39-50. http://dx.doi.org/10.1016/j.jhydrol.2012.12.031.

Brown, A. E., Zhang, L., McMahon, T. A., Western, A. W., & Vertessy, R. A. (2005). A review of paired catchment studies for determining changes in water yield resulting from alterations in vegetation. *Journal of Hydrology*, *310*(1-4), 28-61. http://dx.doi.org/10.1016/j.jhydrol.2004.12.010.

Carlier, C., Wirth, S. B., Cochand, F., Hunkeler, D., & Brunner, P. (2018). Geology controls streamflow dynamics. *Journal of Hydrology*, *566*, 756-769. http://dx.doi.org/10.1016/j.jhydrol.2018.08.069.

Castro, L., & Carvajal, Y. (2010). Análisis de tendencia y homogeneidad de series climatologicas. *Ingeniería de Recursos Naturales y del Ambiente*, *9*, 15-25.

Célleri, R., & Feyen, J. (2009). The hydrology of tropical andean ecosystems: importance, knowledge status, and perspectives. *Mountain Research and Development*, *29*(4), 350-355. http://dx.doi.org/10.1659/mrd.00007.

Cheng, L., Yaeger, M., Viglione, A., Coopersmith, E., Ye, S., & Sivapalan, M. (2012). Exploring the physical controls of regional patterns of flow duration curves – Part 1: insights from statistical analyses. *Hydrology and Earth System Sciences*, *16*(11), 4435-4446. http://dx.doi.org/10.5194/hess-16-4435-2012.

Chiew, F., & Siriwardena, L. (2005). *TREND trend/change detection software: user guide* (pp. 29). Melbourne: Catchment Modelling Toolkit.

Coopersmith, E., Yaeger, M. A., Ye, S., Cheng, L., & Sivapalan, M. (2012). Exploring the physical controls of regional patterns of flow duration curves – Part 3: a catchment classification system based on regime curve indicators. *Hydrology and Earth System Sciences*, *16*(11), 4467-4482. http://dx.doi.org/10.5194/hess-16-4467-2012.

Crespo, P. J., Feyen, J., Buytaert, W., Bücker, A., Breuer, L., Frede, H.-G., & Ramírez, M. (2011). Identifying controls of the rainfall–runoff response of small catchments in the tropical Andes (Ecuador). *Journal of Hydrology*, *407*(1-4), 164-174. http://dx.doi.org/10.1016/j.jhydrol.2011.07.021.

Davis, J. C. (1973). *Statistics and data analysis in geology*. New York: Wiley.

Durán-Quesada, A. M., Reboita, M., & Gimeno, L. (2012). Precipitation in tropical America and the associated sources of moisture : a short review. *Hydrological Sciences Journal*, *57*(4), 612-624. http://dx.doi.org/10.1080/02626667.2012.673723.

Eagleson, P. S. (1970). *Dynamic hydrology*. New York: McGraw Hill.

Galbraith, J. I., Moustaki, I., Bartholomew, D. J., & Steele, F. (2002). *The analysis and interpretation of multivariate data for social scientists*. Boca Raton: Chapman & Hall/CRC. http://dx.doi.org/10.1201/9781420057454.

Garreaud, R. D. (2009). The Andes climate and weather. *Advances in Geosciences*, *22*, 3-11. http://dx.doi.org/10.5194/adgeo-22-3-2009.

Gleeson, T., Wang-Erlandsson, L., Porkka, M., Zipper, S. C., Jaramillo, F., Gerten, D., Fetzer, I., Cornell, S. E., Piemontese, L., Gordon, L. J., Rockström, J., Oki, T., Sivapalan, M., Wada, Y., Brauman, K. A., Flörke, M., Bierkens, M. F. P., Lehner, B., Keys, P., Kummu, M., Wagener, T., Dadson, S., Troy, T. J., Steffen, W., Falkenmark, M., & Famiglietti, J. S. (2020). Illuminating water cycle modifications and Earth system resilience in the Anthropocene. *Water Resources Research*, *56*(4), 1. http://dx.doi.org/10.1029/2019WR024957.

Guajardo, P., & Granados, R. (2010). *Pruebas de homogeneidad de varianza a series de precipitación y temperatura de estaciones del estado de Veracruz y áreas aledañas* (pp. 1-10). Mexico: Organización Mexicana de Meteorólogos.

Guzmán, D., Ruíz, J. F., & Cadena, M. (2014). *Regionalización de Colombia según la estacionalidad de la precipitación media mensual, a través análisis de componentes principales (ACP)*. Bogotá: Instituto de Hidrología, Meteorología y Estudios Ambientales. Retrieved in 2019, August 31, from http://journal.stainkudus.ac.id/index.php/equilibrium/article/view/1268/1127

Hoyos, I., Cañón, F. D. J., & Martínez, B. J. A. (2018). Moisture origin and transport processes in Colombia, northern South America. *Climate Dynamics*, *50*(3-4), 971-990. http://dx.doi.org/10.1007/s00382-017-3653-6.

Hurtado, A., & Mesa, Ó. (2015). Climate change and space-time variability of the precipitation in Colombia. *Revista EIA*, *12*(12), 131-150.

Ilstedt, U., Bargués Tobella, A., Bazié, H. R., Bayala, J., Verbeeten, E., Nyberg, G., Sanou, J., Benegas, L., Murdiyarso, D., Laudon, H., Sheil, D., & Malmer, A. (2016). Intermediate tree cover can maximize groundwater recharge in the seasonally dry tropics. *Scientific Reports*, *6*(1), 21930. PMid:26908158. http://dx.doi.org/10.1038/srep21930.

Instituto de Hidrología, Meteorología y Estudios Ambientales – IDEAM. (2014) *Mapa de coberturas de la tierra: metodología CORINE land cover adaptada para Colombia: período 2010-2012: escala 1:100.000*. Bogotá: IDEAM. Retrieved in 2019, August 31, from http://sirh.ideam.gov.co

Instituto de Investigaciones Ambientales del Pacífico – IIAP. (2013). *Plan estrategico de la macrocuenca del pacifico colombiano*. Bogotá: IIAP.

Instituto Geográfico Agustín Codazzi – IGAC. (2007). *Mapa de suelos de Colombia, escala 1:100.000*. Bogotá.

Intergovernmental Panel on Climate Change – IPCC. (2014a). *Climate change 2014: impacts, adaptation and vulnerability: glossary*. Geneva: IPCC. Retrieved in 2019, August 31, from www.ipcc.ch/report/ar5/wg2/

Intergovernmental Panel on Climate Change – IPCC. (2014b). *Climate change 2014: impacts, adaptation and vulnerability: technical summary*. Geneva: IPCC. Retrieved in 2019, August 15, from www.ipcc.ch/pdf/assessment-report/ar5/ wg2/WGIIAR5-TS_FINAL.pdf

Jackson, J. E. (1991). *A user's guide to principal components* (1st ed.). Hoboken: John Wiley & Sons. http://dx.doi.org/10.1002/0471725331.

Japan Aerospace Exploration Agency - JAXA. (2017). *ALOS (Advanced Land Observing Satellite) PALSAR*. Retrieved in 2018, January 20, from https://vertex.daac.asf.alaska.edu.

Le Maitre, D. C., Scott, D. F., & Colvin, C. (1999). A review of information on interactions between vegetation and groundwater. *Water S.A.*, *25*(2), 137-152.

Legendre, P., & Legendre, L. (2003). *Numerical ecology* (3rd ed.). Amsterdam: Elsevier.

Leyva, P. (1993). *Colombia Pacifico Tomo II*. Bogotá D.C.: Fondo para la Protección del Medio Ambiente José Celestino Mutis.

Masiokas, M. H., Cara, L., Villalba, R., Pitte, P., Luckman, B. H., Toum, E., Christie, D. A., Le Quesne, C., & Mauget, S. (2019). Streamfow variations across the Andes (18°–55°S) during the instrumental era. *Scientific Reports*, *9*(1), 178779. PMid:31784550. http://dx.doi.org/10.1038/s41598-019-53981-x.

Meixner, T., Manning, A. H., Stonestrom, D. A., Allen, D. M., Ajami, H., Blasch, K. W., Brookfield, A. E., Castro, C. L., Clark, J. F., Gochis, D. J., Flint, A. L., Neff, K. L., Niraula, R., Rodell, M., Scanlon, B. R., Singha, K., & Walvoord, M. A. (2016). Implications of projected climate change for groundwater recharge in the western United States. *Journal of Hydrology*, *534*, 124-138. http://dx.doi.org/10.1016/j.jhydrol.2015.12.027.

Mosquera Machado, S. D. C. (2006). Análisis multirriesgos de los desastres naturales del Chocó, durante las tres últimas décadas: 1970-2000. *Entorno Geográfico*, *4*, 1-9.

Ng, G. H. C., McLaughlin, D., Entekhabi, D., & Scanlon, B. R. (2010). Probabilistic analysis of the effects of climate change on groundwater recharge. *Water Resources Research*, *46*(7), 1-18. http://dx.doi.org/10.1029/2009WR007904.

Noori, R., Sabahi, M. S., Karbassi, A. R., Baghvand, A., & Taati Zadeh, H. (2010). Multivariate statistical analysis of surface water quality based on correlations and variations in the data set. *Desalination*, *260*(1-3), 129-136. http://dx.doi.org/10.1016/j.desal.2010.04.053.

Pardo, C., & Alfonso, W. (2018). Climate change in Colombia achieving sustainable development from society. *International Journal of Climate Change Strategies and Management*, *10*(4), 632-652.

Peña-Arancibia, J. L., Bruijnzeel, L. A., Mulligan, M., & van Dijk, A. I. J. M. (2019). Forests as 'sponges' and 'pumps': assessing the impact of deforestation on dry-season flows across the tropics. *Journal of Hydrology*, *574*, 946-963. http://dx.doi.org/10.1016/j.jhydrol.2019.04.064.

Ponette-González, A. G., Marín-Spiotta, E., Brauman, K. A., Farley, K. A., Weathers, K. C., & Young, K. R. (2014). Hydrologic Connectivity in the high elevation tropics: heterogeneous responses to land change. *Bioscience Journal*, *64*(2), 92-104. http://dx.doi.org/10.1093/biosci/bit013.

Poveda Jaramillo, G. (2004). La hidroclimatología de Colombia: una síntesis desde la escala inter-decadal hasta la escala diurna. *Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales*, *28*(107), 201-222.

Poveda Jaramillo, G., & Mesa, O. J. (2000). On the existence of Lloró (the rainiest locality on Earth): enhanced ocean-atmosphere-land interaction by a low-level jet. *Geophysical Research Letters*, *27*(11), 1675-1678. http://dx.doi.org/10.1029/1999GL006091.

Poveda Jaramillo, G., Jaramillo, L., & Vallejo, L. F. (2014). Seasonal precipitation patterns along pathways of South American low-level jets and aerial rivers. *Water Resources Research*, *50*(1), 98-118. http://dx.doi.org/10.1002/2013WR014087.

Price, K. (2011). Effects of watershed topography, soils, land use, and climate on baseflow hydrology in humid regions: a review. *Progress in Physical Geography*, *35*(4), 465-492. http://dx.doi.org/10.1177/0309133311402714.

R Core Team. (2018). *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Retrieved in 2019, August 31, from https://www.r-project.org

Richman, M. B. (1981). Obliquely rotated Principal Components: an improved methodological map typing technique? *American Meteorological Society*, *20*(10), 1145-1159. http://dx.doi.org/10.1175/1520-0450(1981)020<1145:ORPCAI>2.0.CO;2.

Rueda, Ó. A., & Poveda Jaramillo, G. (2006). Spatial and temporal variability of the Choco jet stream and its effect on the hydroclimatology of the Colombian pacific. *Meteorologia Colombiana*, (10), 132-145. Retrieved in 2019, August 31, from http://ciencias.bogota.unal.edu.co/fileadmin/content/geociencias/revista_meteorologia_colombiana/numero10/10_12.pdf

Sharma, S. K., Gajbhiye, S., & Tignath, S. (2015). Application of principal component analysis in grouping geomorphic parameters of a watershed for hydrological modeling. *Applied Water Science*, *5*(1), 89-96. http://dx.doi.org/10.1007/s13201-014-0170-1.

Singh, P. K., Kumar, V., Purohit, R. C., Kothari, M., & Dashora, P. K. (2009). Application of principal component analysis in grouping geomorphic parameters for hydrologic modeling. *Water Resources Management*, *23*(2), 325-339. http://dx.doi.org/10.1007/s11269-008-9277-1.

Smakhtin, V. U. (2001). Low flow hydrology: a review. *Journal of Hydrology*, *240*(3-4), 147-186. http://dx.doi.org/10.1016/S0022-1694(00)00340-1.

Sophocleous, M. (2002). Interactions between groundwater and surface water: the state of the science. *Hydrogeology Journal*, *10*(1), 52-67. http://dx.doi.org/10.1007/s10040-001-0170-8.

Strauch, A. M. (2013). Interactions between soil, rainfall, and wildlife drive surface water quality across a savanna ecosystem. *Ecohydrology*, *6*(1), 94-103. http://dx.doi.org/10.1002/eco.1246.

Strauch, A. M., MacKenzie, R. A., Giardina, C. P., & Bruland, G. L. (2015). Climate driven changes to rainfall and streamflow patterns in a model tropical island hydrological system. *Journal of Hydrology*, *523*, 160-169. http://dx.doi.org/10.1016/j.jhydrol.2015.01.045.

Urrea, V., Ochoa, A., & Mesa, O. (2019). Seasonality of rainfall in Colombia. *Water Resources Research*, *55*(5), 4149-4162. http://dx.doi.org/10.1029/2018WR023316.

Vélez, A. C., & Aguirre, R. N. (2016). Influencia del río atrato en el golfo de Urabá durante el holoceno tardío, mar caribe colombiano. *Boletin de Investigaciones Marinas y Costeras*, *45*(1), 73-97.

Westra, S., Brown, C., Lall, U., & Sharma, A. (2007). Modeling multivariable hydrological series: principal component analysis or

independent component analysis? *Water Resources Research*, *43*(6). http://dx.doi.org/10.1029/2006WR005617.

Zea, J. A., León, G. E., & Eslava, J. A. (2000). Influencia de sistemas extratropicales en Colombia. *Meteorología Colombiana*, *1*, 31-40.

Authors contributions

Sebastián Balbín Betancur: Contribution: Conceptualization, Data analysis, Writing-original draft, Writing-review & editing.

Didier Gastmans: Funding acquisition, Supervision, Data analysis, Writing-original draft, Writing-review & editing.

Katherine Vásquez Vásquez: Data analysis and Investigation.

Lucas Vituri Santarosa: Data analysis and Investigation.

Vinícius dos Santos: Data analysis and Investigation.

Roberto Eduardo Kirchheim: Data analysis, Writing-original draft, Writing-review & editing.