An Acad Bras Cienc (2022) 94(4): e20201000 DOI 10.1590/0001-3765202220201000
Anais da Academia Brasileira de Ciências | *Annals of the Brazilian Academy of Sciences*
Printed ISSN 0001-3765 I Online ISSN 1678-2690
www.scielo.br/aabc | www.fb.com/aabcjournal

GEOSCIENCES

Techniques for monthly rainfall regionalization in southwestern Colombia

TERESITA CANCHALA, CAMILO OCAMPO-MARULANDA, WILFREDO ALFONSO-MORALES, YESID CARVAJAL-ESCOBAR, WILMAR L. CERÓN & EDUARDO CAICEDO-BRAVO

Abstract: The knowledge of rainfall regimes is a relevant requirement for many activities such as water resources planning, risk management, agriculture activities management, and other hydrologic applications. The present study has consisted of validating four techniques (one linear, one non-linear, and two hybrids) that allow identifying homogenous regions. We take the monthly rainfall in the Southwestern Colombia (Nariño), an area of 33,268 km^2 characterized by complex topography and local factors that can influence the rainfall behavior, to test all techniques. The results showed overall the best performance for the approach related to non-linear principal component analysis and self-organizing map. However, in all mainly prevail two regions: the Andean Region and Pacific Region with a bimodal and unimodal regime, respectively. The bimodal one dominates over the Andes mountains range and the unimodal one the coastal zone. The application of non-linear approaches provided a better understanding of the seasonality of rainfall, and the results may be useful for water resource management.

Key words: rainfall, regionalization, Principal Component Analysis (PCA), Hierarchical Clustering Analysis (HCA), Non-Linear Component Analysis (NLPCA), Self-Organizing Map (SOM).

INTRODUCTION

Rainfall is one of the meteorological variables that deserve further study in tropical areas due to it is the primary water source for agriculture activities, varies over space and time, and can be a good indicator of climatic variability. The physics of the rainfall formation process is complex and has no accuracy about its formation and forecasting (Ramirez et al. 2005). The need to understand the distribution, intensity, and frequency of rainfall in different regions, has promoted the delineation of homogeneous rainfall zones as a fundamental step in their analysis (Mannan et al. 2018) through cluster analysis. Cluster analysis or "clustering" is a technique used for grouping data points, objects, or vectors with similar properties and features; nevertheless, the clusters are different from each other (Gocic & Trajkovic 2014).

The cluster analysis on rainfall is a complex process due to dynamism and non-linearity at spatial and temporal scales. For this reason, different methods have been used such as (i) correlation analysis (Shen 2019, Seo et al. 2019), (ii)Principal Component Analysis (PCA) (Asong et al. 2015, Guzmán 2016), (iii) spectral analysis (Magallanes et al. 2015), (iv) Hierarchical Cluster Analysis (HCA) (Santos et al. 2015), (v) PCA in combination with HCA (Satyanarayana and Srinivas 2011, Gocic & Trajkovic 2014, Fazel et al. 2018, Raziei 2018), and the over the last decades (vi) Kohonen Self-Organizing Maps (SOM) (Lin & Chen 2006, Farsadnia et al. 2014, Rau et al. 2017, Mannan et al. 2018, Markonis & Strnad 2019).

These methods allow the rainfall regionalization, through the identification of clusters with similar features of rainfall, to facilitate the research of regional climate variables, to improve our understanding about aspects linked to climatic variability (Wong et al. 2016), the dynamics of water balance at several scales (Rau et al. 2017), to bring support in the decision making in water resource management (Guerrero et al. 2006), to the planning of agriculture (Satyanarayana & Srinivas 2008), and for rainfall forecasting (Asong et al. 2015).

According to Srinivas (2013), rainfall regionalization is a relevant process in places with high spatial and temporal variability to obtain an adequate analysis of rainfall patterns and temporal scales. Colombia requires a detailed analysis of its hydro-climatology due to its complex orographic features imposed by the Andes Mountain Range (Garreaud et al. 2009, Poveda et al. 2020), and its hydroclimatic conditions (Poveda et al. 2011, Poveda 2004). The large-scale hydrometeorological phenomena with the most influence over Colombian territory include El Niño-Southern Oscillation (ENSO) (Poveda et al. 2006, Puertas & Carvajal 2008, Poveda et al. 2011), the Atlantic Multidecadal Oscillation (AMO) (Cerón et al. 2020), and the Pacific Decadal Oscillation (PDO) (Poveda 2004, Cerón et al. 2020).

Furthermore, the rainfall variability in Colombia is strongly modulated by inter-seasonal mechanisms like the Intertropical Convergence Zone (ITCZ) (Poveda & Mesa 1997, Poveda et al. 2006), the Choco low-level jet (CJ) (Hoyos et al. 2018, Yepes et al. 2019, Serna et al. 2018), the Caribbean low-level jet (Arias et al. 2015, Serna et al. 2018), the Panama jet (Yepes et al. 2019), Easterly Waves (Dominguez et al. 2020), Mesoscale Convective System (Jaramillo et al. 2017), and the advection from the Amazon and Orinoco Basins (Poveda et al. 2006).

Considering the aforementioned mechanisms is relevant to highlight the ITCZ, and the CJ, which exerts a strong modulation over the Colombia hydroclimatology. According to Poveda & Mesa (1997), and Poveda et al. (2006), the ITCZ meridional migration represents the principal factor driving the annual cycle across Colombia. The double passage of ITCZ during the austral and boreal summer over Colombia establish a bimodal annual cycle in western and central Colombia, featuring two wet seasons from March–May (MAM) and September–November (SON), and two dry seasons from December–February (DJF) and June–August (JJA). Also noteworthy is the CJ, a southwesterly circulation pattern acting over the eastern tropical Pacific, where moisture is form transported from the Chilean coast to the Colombian Pacific coast (Hoyos et al. 2018, Yepes et al. 2019). Upon entering the continent, the CJ is lifted by the orography of the Western Cordillera of the Colombian Andes, interacts with the trade winds, and contributes to enhancing deep convection (Jaramillo et al. 2017, Bedoya-Soto et al. 2019). CJ exhibits an annual cycle, with strengthening during SON with speeds varying between 5 and 7 $m.s^{-1}$, and weakening during MAM, which is no more than 3 $m.s^{-1}$. Thus, CJ modulates the wet season over western and central Colombia, mainly during SON (Poveda & Mesa 2000, Serna et al. 2018, Cerón et al. 2020).

Several studies in Colombia show the implementation of rainfall regionalization techniques with different purposes such as the determination of the incidence of the El Niño and La Niña phenomenon on crop productivity (Barrios, unpublished data), the establishment of modes of climatic variability (Carmona & Poveda 2012, Canchala et al. 2020), the identification of seasonal and temporal monthly rainfall variability (Guzmán et al. 2014, Carvajal & Segura 2004, Estupiñan 2016), and the estimation of missing data (Samuel et al. 2011). However,

there are no specific studies about monthly rainfall in Southwestern Colombia (Nariño), despite this area has a complex topography and is located in a geostrategic position between the Colombian-Ecuadorian border, the Tropical Pacific Ocean and the Andes Mountain Range of South America.

Since rainfall are intrinsically dynamic and stochastic, we raise two objectives. First, to compare four techniques for monthly rainfall regionalization in Nariño for the period 1983-2016: a representative traditional linear approach PCA+HCA, a non-linear combination of Non-Linear Principal Component Analysis–NLPCA+SOM, and two-hybrid ones (PCA+SOM and NLPCA+HCA). Second, to identify spatial distribution and seasonality of the monthly rainfall. The section below presents the details of the study area and data, followed by the description of the applied methods and the results and discussion. Finally, the conclusions and future directions are presented.

STUDY AREA AND DATA

Study area

The study area lies in Southwestern Colombia (Nariño) in the south of Colombian Biogeographic Choco, (Figure 1). This region covers an area of 33,268 km^2, is crossed by the Andes Mountain, and is near to the Pacific Ocean.

Rainfall datasets

Series of monthly rainfall with 34 years of observation between 1983 and 2016 corresponding to forty-five rainfall-gauge stations located in different zones in Nariño, provided by Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) of Colombia, are part of this study (Figure 1). Descriptive statistical details of each rainfall gauge station are available in Table I registered by Canchala et al. (2019).

Pre-processing

Filling missing data

The estimation of these missing data was possible through Artificial Neural Networks (ANN), according to the method proposed by Scholz et al. (2005). This method uses the inverse model of the NLPCA technique using a reconstruction function Φ_{gen} which is performed by a feed-forward network. The goal of this function is generating data $\hat{x}$ that approximates the target data X by minimizing the squared error $X-\hat{X}^2$. The work from Canchala et al. (2019) details this procedure for the same period's gauge-stations. After completing this procedure for each series, the Root Mean Square Error (RMSE) helps to evaluate the accuracy of the estimate method

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(e_i-o_i\right)^2} \quad \text{Eq. 1}$$

where e is the estimated value, o indicates the observed value, and n refers to sample size.

Data consistency analysis

Using Pearson's correlation coefficient (CC), we obtained a consistency analysis of the 45 monthly rainfall time-series datasets. The CC was used to quantify the correspondence between gauge stations and identify gauge stations anomalous. CC varies from -1 to 1 being the maximum the perfect positive correlation. For the present research work, the statistical significance was determined by the Student's t-test for a confidence value $\alpha=0.05$.

Transformation of data

Before to cluster analysis through the different methodologies, the rainfall observation r_{ij}

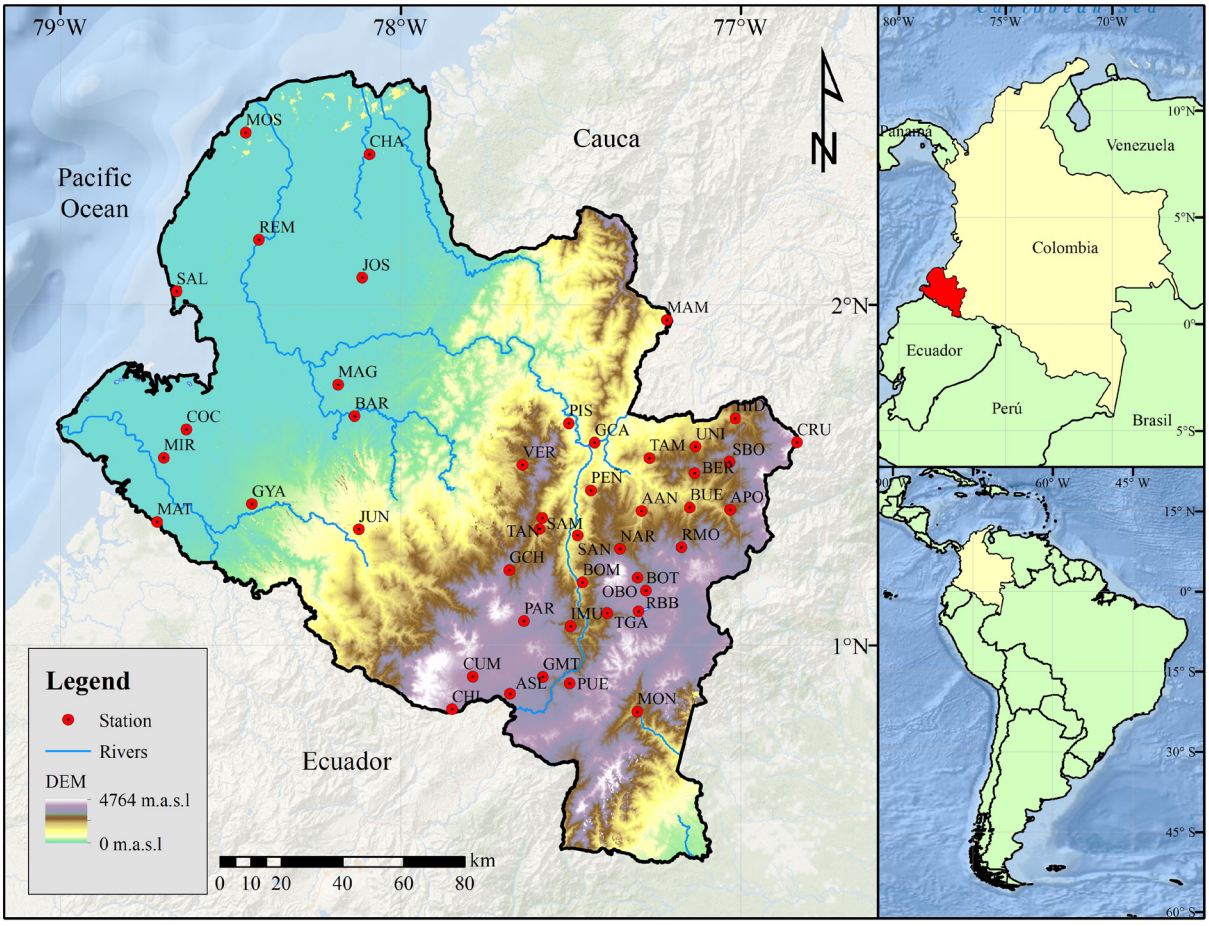

Figure 1. The geographic location of the study area and distribution of rainfall gauge stations.

of station *j* at a time *i* is transformed into a standard score using:

$$Z_{ij} = \frac{r_{ij} - \overline{r_j}}{s_j}, \quad \text{Eq. 2}$$

where $\overline{r_j}$ and s_j are the average and standard deviation of rainfall data of station *j*, respectively.

MATERIALS AND METHODS

Methods applied

Principal component analysis

PCA is a widely used linear statistical procedure that seeks to reduce the number of components from a climatological rainfall database. These components correspond to the uncorrelated orthogonal base (Principal Components – PCs) (Miró et al. 2017). PCA calculates the eigenvectors and eigenvalues from the correlation matrix. The eigenvectors provide information concerning the weight that the original data have in the recent-formed components, and the eigenvalues give the value of the explained variance of each new variable (Martins et al. 2012). Besides, PCA provides the PCs sorted in descending order of importance.

In this work, the PCs (u) are linear combinations of the standardized rainfall $F' = F - \overline{F}$, where the data matrix includes N columns (stations), and M rows (monthly rainfall values).

$$u = E^T F' \quad \text{Eq. 3}$$

The columns of the orthogonal matrix E are the eigenvectors, well-known as empirical orthogonal functions. The superscript T indicates the transpose operation. Every PC depicts a percentage of the total variance in F that is proportional to its eigenvalue (Tošić et al. 2016). The Kaiser Meyer-Olkin (KMO) statistic helps to test the quality of the obtained PCs, which compares the magnitude from both the correlation and the partial correlation coefficients. Its index varies between 0 and 1, and the assumption of multicollinearity is accepted when KMO > 0.5 (Sheskin 2003). The PCs with eigenvalues beyond one were selected, as suggested by Jolliffe (1986) and Preisendorfer (1988).

Non-linear principal component analysis

NLPCA is developed as a non-linear generalization of the classic PCA method (Hsieh 2001), considering the PCA limitations when it involves non-linear processes (Miró et al. 2017). NLPCA operates by training a feed-forward neural network to perform the identity mapping, where the network inputs and outputs are the same. The neural network contains an internal "bottleneck" layer that allows generating a compact representation of the input data. This method successfully reduces the dimensionality and create a feature space map similar to the actual distribution of the underlying system parameters (Kramer 1991). In this method, the PCs turn from the right lines to curves; therefore, the subspace in the original data-space also is depicted in a curved mode (Miró et al. 2017). Details about the Neural Network model for calculating NLPCA is available in Hsieh (2001) and Scholz et al. (2008).

Hierarchical cluster analysis

For clustering purposes, we use the HCA with Euclidean distance measure for the observations and Ward's method for the linkage rule. According to Darand & Daneshvar (2014) and Hershey et al. (2010), this fusion results in the most distinctive clusters. Therefore, for the first stage of the clustering process, which is the estimation of similarity (or dissimilarity), we used Euclidean distance. The Euclidean distance between two elements $X=[X_1, X_2, ..., X_n]$ and $Y=[Y_1, Y_2, ..., Y_n]$ is defined by

$$d_{xy} = \sqrt{(X_1 - Y_1)^2 + (X_2 - Y_2)^2 + ... + (X_n - Y_n)^2} = \sqrt{\sum_{i=1}^{n} (X_i - Y_i)^2} \quad \text{Eq. 4}$$

where, X_i and Y_i are the elements to be compared, which in this study are the standardized rainfall. Euclidean distance is the most commonly used measure, although many other distance measures exist (Gong & Richman 1995).

The second step is to define the clustering method. This study chose Ward's method because it has been the clustering technique most used in climate regionalization (Fazel et al. 2018). Overall, it outperforms other algorithms in terms of segregation, providing relatively dense groups with minimums within-group variance. Ward's method recognizes the minimum variance within groups, joining elements with a minimal sum of squares between them (Hervada-Sala & Jarauta-Bragulat 2004, Santos et al. 2015).

Self-organizing maps

SOM is a non-linear method that allows the clustering, visualization, and abstraction of complex data. SOM is a type of ANN trained using unsupervised learning. SOM approximates the probability density function of the input data by an unsupervised learning algorithm, with properties of neighborhood preservation and local resolution of the input space proportional

to the data distribution (Kohonen 2001, 1982). SOM is composed of two layers: an input layer composite by a set of nodes, and an output layer composite by nodes ordered in a two-dimensional grid (Figure 2) (Hsu & Li 2010). Each node in the input layer is joined to all the nodes in the output layer by synaptic links. Each output node has a weight vector ***W*** linked with the input data, which establishes a relationship between the feature vector and the cluster of feature vectors (Farsadnia et al. 2014). Due to its unique ability to locate a pattern even in complicated structures and the non-linear characteristics of climate, in the last years, SOM has become in a clustering method of hydroclimatic variables with promising results (Hsu & Li 2010, Nourani et al. 2012, Farsadnia et al. 2014, Agarwal et al. 2016, Mannan et al. 2018, Wang & Yin 2019).

Classification and regionalization process

Classification and regionalization process follows four techniques i) linear approach (PCA+HCA), ii) non-linear approach (NLPCA+SOM), iii), and iv) hybrid approaches (PCA+SOM and NLPCA+HCA), describes in next sections. The methodology is composed of three steps summarized in Figure 3. The first step is the pre-processing, which includes the estimation of missing data and the consistency analysis using Pearson's correlation coefficients to identify atypical gauge stations and discard the anomalous dataset. The second step corresponds to the regionalization process, which includes dimensional reduction (PCA or NLPCA), clustering (HCA or SOM), validation, and selection of the best method. The last one is the spatial coherence verification and detailed characterization of rainfall patterns for each identified region.

Linear approach (PCA+HCA)

PCA, in association with HCA, is the most commonly used technique to identify homogeneous rainfall regions (Raziei et al. 2008, Darand & Daneshvar 2014, Gocic & Trajkovic 2014) having a notable performance in the regionalization process. Both methods are complementary: PCA can reduce the dimensionality of data, and HCA classifies sub-regions with similar rainfall features (Fazel et al. 2018). To identify homogeneous groups in our dataset following this technique, PCA was used to reduce the dimensionality of the dataset following Eq. (3). The KMO statistic helps establish the quality of the retained PCs, which are the inputs for the HCA. In the HCA application, the Euclidean distance (Eq. 4) measures the observations, and Ward's method is the linkage rule.

Non-linear approach (NLPCA+SOM)

Taking into account the advantages of non-linear methods such as the increased power of components coming from NLPCA (Hsieh 2001, Scholz & Vigário 2002, Miró et al. 2017) and the regardless the shape of SOM (Kohonen 2001, Farsadnia et al. 2014, Wang & Yin 2019), we considered combining NLPCA and SOM. NLPCA was used to reduce the temporal dimensionality of the dataset and SOM to capture a profile of the homogeneous areas and to obtain a classification of the gauge stations. The numbers of estimated NLPCs are equivalent according to the number of PCs established in a linear approach to be able to make comparisons, and these NLPCs are the inputs of SOM. We chose 25 outputs nodes map (grid of 5 x 5 cells) to improve the visualization. Since it occurs a random initialization of each SOM, the results can be different. Therefore, 10 SOMs were trained for each realization, choosing the one with the best result according to the performance metrics.

Hybrid approach (PCA+SOM and NLPCA+HCA)

In the field of hydro-climatology, some studies have applied hybrid techniques that use a

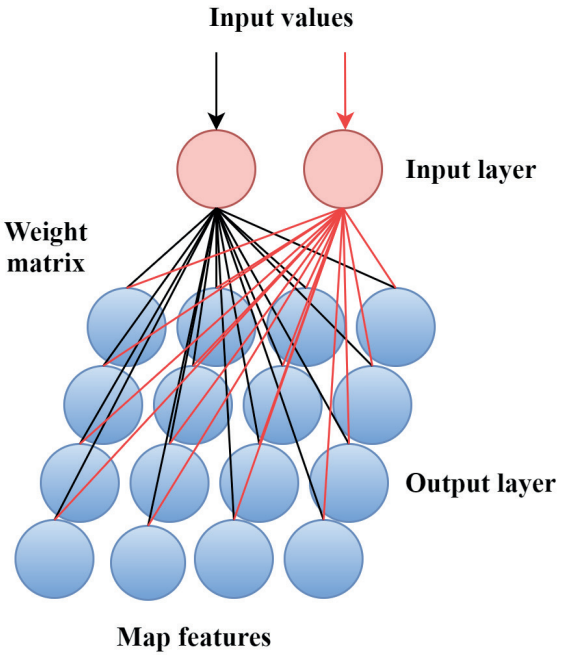

Figure 2. SOM two-level architecture.

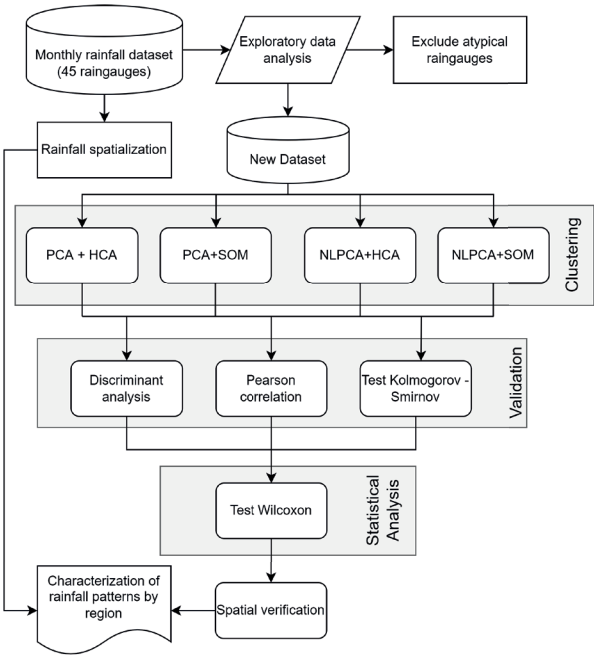

Figure 3. Flowchart of methodology.

set of methods (Linear and non-linear) as a way to improve results, due to these are complementary. Miró et al. (2017) reported, in the estimation of missing data, that the non-linear computing (related to ANN or hybrid) is more efficient than linear or traditional inferences, Chen & Hong (2012) showed that PCA+SOM+L-moments is a robust and effective technique for regional rainfall frequency analysis. Chen et al. (2011) demonstrated that combining PCA and SOM is useful for the regionalization of design hyetographs. In this research, we mixed linear and non-linear methods to get two techniques: PCA+SOM and NLPCA+HCA. The first step in each approach corresponds to dimensional reduction (PCA or NLPCA), while the second one is the clustering (SOM or HCA) to build homogeneous groups from the dataset. The results from the first step are the inputs for the second one.

Validation and performance criteria

The statistical criteria applied to evaluate the whole procedures were: the PCA and CC, the Discriminant Analysis (DA), and the Kolmogorov-Smirnov Test (K-S).

Principal component analysis and Pearson's Correlation

Before calculating the correlation analysis, PCA was used to reduce the size of the time series of rainfall in each obtained group. For this purpose, the first orthogonal PCs were estimated for each region to explain most of the variability of the original variables with minimum loss of information (Jolliffe & Cadima 2016). To examine the dependence of the regional rainfall on each other, we calculated CC of the first PCs per region. Low correlation coefficients indicate that the regional variations in rainfall are from afar independent of each other (Li-Juan et al. 2009), while values close to one indicate homogeneity between regions. The significance of the correlations was identified by the Student's t-test using the p-value for a 95% confidence value ($\alpha = 0.05$).

Discriminant analysis

DA is a supervised learning method that processes a dataset with one classification variable and one or more quantitative variables to describe them (Toth 2013). The quantitive variables are also known as discriminants. The DA is probabilistic, founded on the presumption that the observation in the class is caused by probability distribution specific to the class $f_k(.)$. If τ_k is the ratio of members of the population that are in class k, Bayes's theorem says that the posterior probability that an observation y belongs to class k is:

$$Pr[y \in class\ j] = \frac{\tau_j f_j(y)}{\sum_{k=1}^{G} \tau_k f_k(y)} \quad \text{Eq. 5}$$

Assigning y to the class to which it has the highest posterior probability of belonging minimizes the expected misclassification rate; this is called the Bayes classifier (Fraley & Raftery 2002). In the present research, the quantitative variables describing each entity are the first PCs or NLPCs of the monthly rainfall. The classes are the cluster identified by the HCA or SOM network.

Kolmogorov-Smirnov test

K-S test helps to evaluate the homogeneity of the identified sub-regions; this is a non-parametric statistical test that deduces whether the subjacent distribution of two datasets is significantly different (Raziei 2018, Shen et al. 2016). In this research, we used the K-S test to deduce whether the monthly rainfall of different regions identified in the cluster analysis is statistically different.

Let $F_1(x)$ and $F_2(x)$ be two cumulative distribution functions, the null hypothesis is:

$$H_o: F_1(x) = F_2(x) \text{ for any } x \quad \text{Eq. 6}$$

The alternative hypothesis is:

$$H_o: F_1(x) < F_2(x) \text{ for any } x \quad \text{Eq. 7}$$

Wilcoxon signed-rank test

We use the Wilcoxon signed-ranks test (Wilcoxon 1992) to establish the best regionalization approach considering the results obtained in the three performance criteria (CC, DA, and K-S). Wilcoxon signed-ranks test is a non-parametric alternative to the paired t-test, which ranks the differences in performances of two classifiers for each dataset, ignoring the signs, and comparing the ranks for positive and negative differences (Demšar 2006).

Spatial verification and analysis of monthly rainfall

We verify the spatial coherence of stations classified in each cluster by the stations' localization on the department of Nariño and assigning different colors and symbols for each one. This analysis allows the establishment of geographic, bioclimatic, and ecosystem coherence of identified regions, a relevant process since the climate and rainfall in Colombia are the result of the interaction of atmospheric, biophysical, and geographic factors (Poveda 2004). The analysis of the annual rainfall cycle was carried out with the identified regions along with their spatial distribution. Finally, to understand the influence of topography on its behavior, a cross-sectional profile of the study area was constructed with data obtained from the Google Earth GIS (Wang et al. 2013).

RESULTS AND DISCUSSION

Overall pre-processing data

The results of the estimation of all stations' missing data are available in Canchala et al. (2019). The best outcome, obtained with [45-44-45] nodes per layer, achieves an RMSE in the imputation of missing data corresponding to 9.8 mm. month^{-1}.

With the estimation of missing data, the respective data consistency analysis was carried out by CC for all observed monthly mean time series of rainfall: The correlation matrix presented (Figure 4) allows us to detect the anomalous stations. The CC between Monopamba (MON) gauge station and the other 44 stations are negatives and show a moderate negative relationship, behavior different from that recorded by the other stations.

Given the behavior of MON, the geographic location, and the annual cycle of monthly rainfall, we isolated it to know more about this gauge station. MON is the only station of Nariño found on the eastern slopes of the Andes Mountain Range, over the Amazon basin (See Figure 1). The annual cycle rainfall showed a unimodal regime (with one dry season and one rainy season) with high rainfall from June to August (JJA) (See Figure 5), coherent with Urrea et al. (2019). They used series of daily rain from the IDEAM data set and found that the unimodal regime occurs in the Amazon region of Colombia. Due to MON's rainfall regime is different from that registered by the other 44 gauge stations, and trying to define the best method of regionalization of the monthly rainfall of Nariño, we excluded the MON gauge station from this analysis. MON is the only station that exists in the Amazon region of Nariño and that during the clustering process would lead to grouping itself in a separate cluster. Therefore, in this study, the following analyzes are performed with 44 gauge stations.

Rainfall classification

We performed a cluster analysis of the monthly rainfall by applying four regionalization techniques (PCA+HCA, PCA+SOM, NLPCA+HCA, and NLPCA+SOM) on the 44 rainfall gauge stations, previously selected. Each one forced to obtain 2, 3, and 4 clusters.

For reduction of dimensions in all approaches were estimated PCs and NLPCs by PCA and NLPCA, respectively. In the PCA, the KMO index calculated for the monthly rainfall series corresponded to 0.97, thus suggesting that the dataset is suitable for PCA. According to Diaz & Morales (2002), KMO Test > 0.6 is tolerable for factor analysis purposes, due to the KMO indicated high multicollinearity among the monthly records of rainfall gauge stations. The PCA results get five PCs with eigenvalues beyond one, which explained 52%, 14%, 4.1%, 2.7%, and 2.5% respectively, i.e., 75.3% of the explained variance. Otherwise, NLPCA was used to reduce the dimensionality of the dataset to five NLPCs, an equal number of components established in the linear approach for comparison. To obtain the NLPCs, we use a network with a [408-200-25-5] topology and maximum iterations set at 7000.

In linear approach, HCA was applied to the five PCs using Ward's method and Euclidean distance to obtain the regionalization of gauge stations, whereas, for the non-linear approach, SOM was applied to the five NLPCs for the same purpose. According to Chen et al. (2011), if the SOM-based clustering result is reasonable and satisfactory, the cluster analysis is accepted. The initial weight vector values in the input and the hidden layers are random values between 0.0 and 1.0, and the learning-rate parameter begins at 0.1 and ends at 0.01. For the hybrid approach, SOM (HCA) receives the five PCs (NLPCs) from performing the clustering process.

Table I present the results obtained with all approaches with the number of gauge stations per each Cluster (C). Figure 6 shows the gauge station allocation considering the regionalization process per technique.

The regionalization of the gauge station in two clusters using four different techniques showed results exactly alike using PCA+HCA,

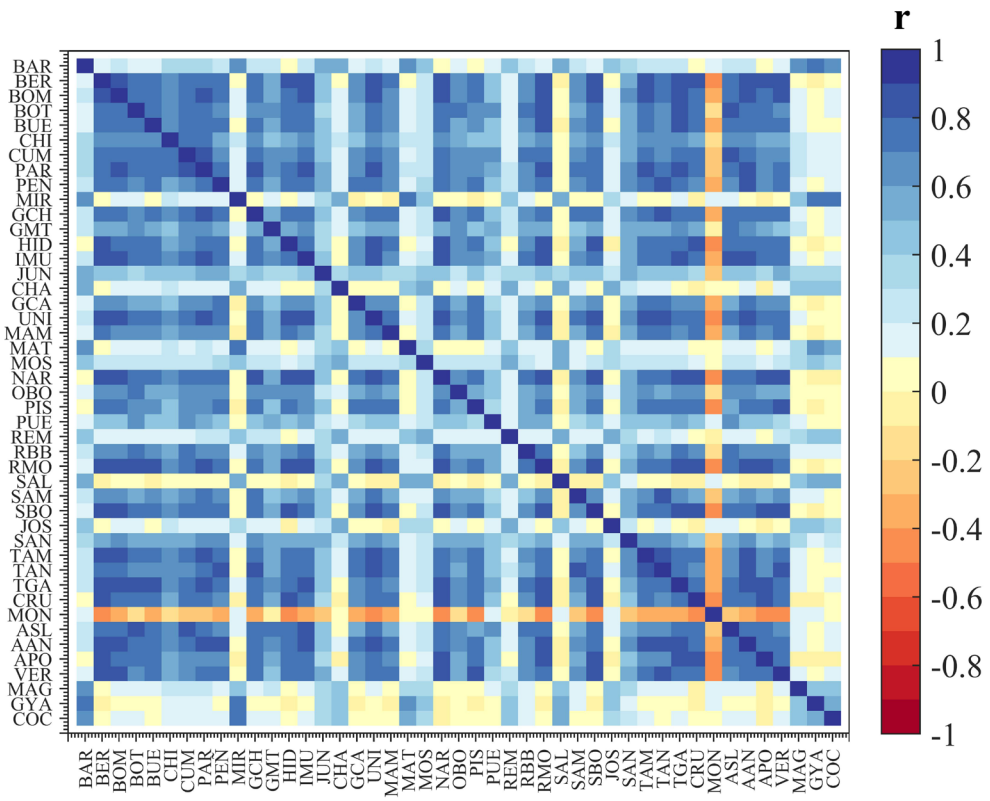

Figure 4. Pearson's correlation coefficients between gauge stations.

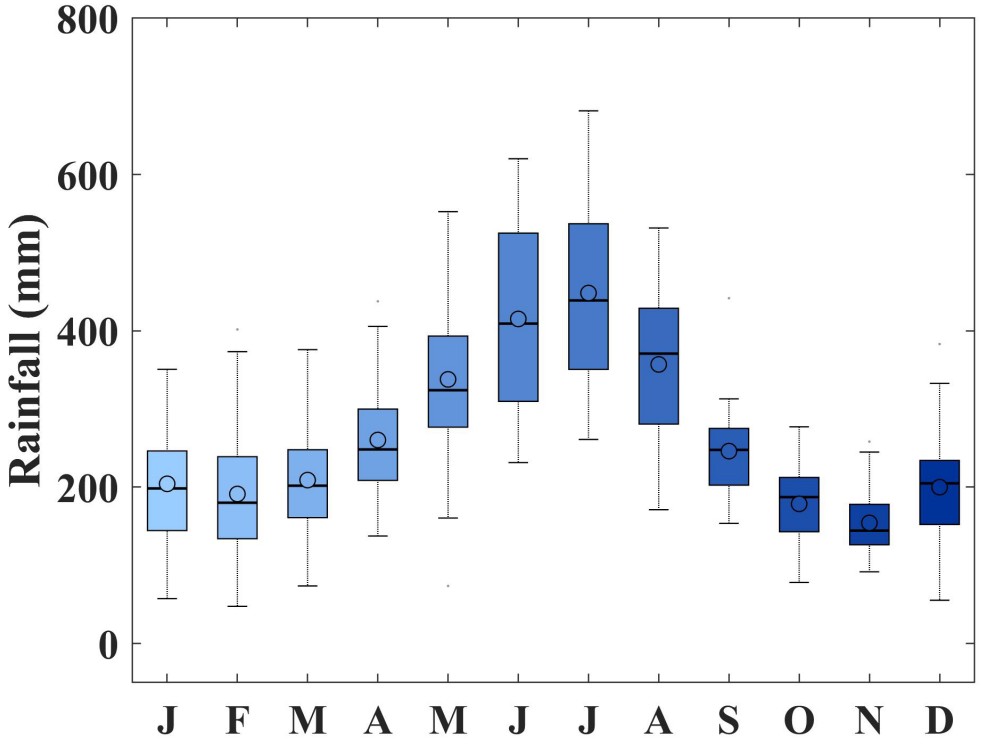

Figure 5. Monthly mean rainfall regime for the period 1983-2016 of MON gauge station. Lower and upper box boundaries are 25th and 75th percentiles, respectively, line inside the box is median, lower and upper error lines are 10th and 90th percentiles, respectively, the hollow point is the mean.

PCA+SOM, and NLPCA+SOM. These techniques located the gauge stations contained in C1 (C2) on the west (east) that corresponds to Pacific coastal (Andes Mountain) of Nariño. This topographic condition can influence the local climate and, therefore, the regionalization process. Meanwhile, the result obtained by NLPCA+HCA showed that C2's stations are over the Andes Mountain and the Pacific coastal plain, and the C1's stations are over the north of the Pacific coastal plain.

The PCA+HCA gets three clusters by splitting the eastern part of Nariño into two, C2 (C3) located over the south (north) of the Andes Mountain of Nariño, and C1's stations are located over Pacific coastal of Nariño. In this clustering process, there is no clear evidence of a clear boundary between the regions C2 and C3. Meanwhile, the results obtained using PCA+SOM and NLPCA+SOM were the same, by splitting the coastal area of Nariño into two, encompassing C1 and C3 the north and south of this area, respectively. Regard to the results obtained by NLPCA+HCA, these were like the described previously, with the difference that in each group identified the number of grouped stations differ (See Table I).

Finally, the clustering of stations in four clusters using the linear, non-linear, and hybrid techniques showed different groupings (Figure 6). Although the number of stations contained in each cluster is different, in the use of non-linear and hybrid methods, the grouped gauge stations located over the mountainous region of Nariño were the same. In contrast, the number of gauge station groups of the Pacific coastal of Nariño corresponded to three, i.e., to the north, center, and south of this region.

Conversely, the PCA+HCA results showed that the grouped gauge stations of the Andes Mountain of Nariño were two but without a clearly defined boundary between the two zones. Here, C1 (C4) includes the north(south) of the Pacific Coastal of Nariño, and C3 (C2) corresponds to the north (south) of the Andes Mountain of Nariño.

Validation

In this research, we applied three performance measurement metrics to evaluate if the clustering process is adequate. Table II shows the performance of regionalization measured through CC between the first PCs of each region for two, three, and four groups. The results

Table I. Results of regionalization by linear, non-linear and hybrid approaches

Number of clusters	Name of cluster	PCA+HCA	PCA+SOM	NLPCA+HCA	NLPCA+SOM
2	C1	11	11	7	11
	C2	33	33	37	33
3	C1	11	5	7	5
	C2	12	33	32	33
	C3	21	6	5	6
4	C1	5	5	5	5
	C2	12	32	32	33
	C3	21	6	2	2
	C4	6	1	5	4

C: Cluster

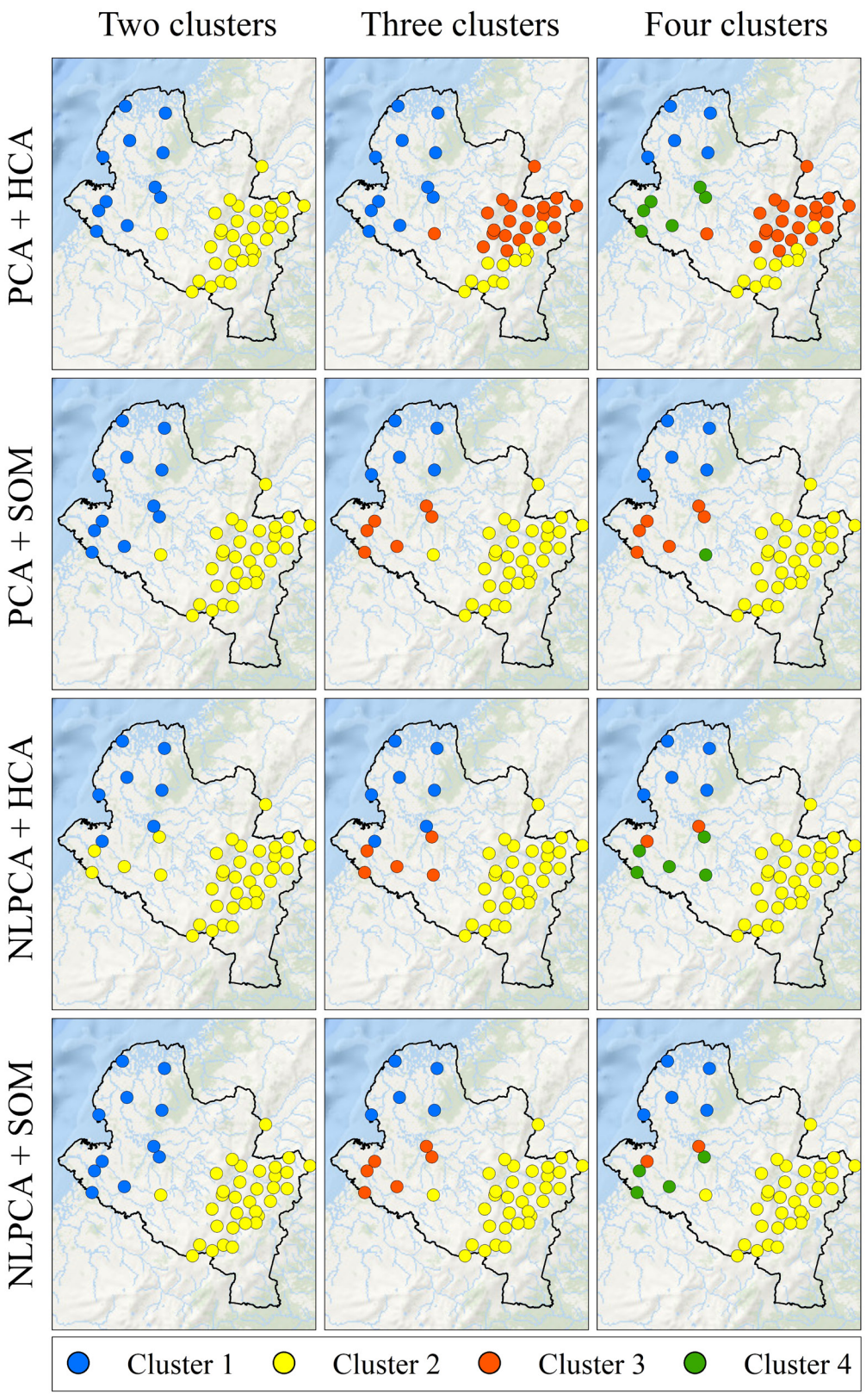

Figure 6. Spatial distribution of gauge stations using linear, non-linear and hybrid approach.

showed that when we compare the rainfall of the clusters located on the Pacific coast and the mountainous area of Nariño, the CC is low, meaning that the rains in these two areas of Nariño are heterogeneous. In contrast, the results for three and four groups showed that correlations between north and south of the Pacific coastal and Andes Mountain of Nariño are high (i.e., the CC between C3 and C1 for three groups using PCA+SOM, NLPCA+HCA, and NLPCA+SOM is 0.69, -0.77, and 0.69, respectively), indicating that the rainfall in the analyzed clusters shares common features. Therefore, for this study, when the number of clusters increases, there is a greater between cluster correlation.

The second performance metric used was the DA, which qualifies the probability of the stations belonging to the assigned cluster with a value from 0 to 1. Table III shows the percentage of stations that presented the maximum likelihood of belonging to the assigned cluster. The results show 100% classification accuracy of gauge stations classified using the non-linear and hybrid approach for two, three, and four groups. In contrast, the linear method's classification results showed that not all stations had the maximum probability for 2, 3, and 4 groups.

Finally, we evaluate the homogeneity of the identified regions using the K-S test to infer whether the monthly rainfall from obtained groups are statistically different, considering that the null hypothesis indicates that the distributions are the same. Table IV shows that almost all comparisons accepted the alternative hypothesis, but C3 and C1 in NLPCA+HCA for four groups configuration not. The measurement of the difference between regions uses the 5% significance level for all techniques.

Due to the performance metrics used are inconclusive, we used the Wilcoxon's rank-sum test; Table V shows their results. The best regionalization techniques, according to these results, are PCA+SOM and NLPCA+SOM (including SOM for the clustering process). Still, the latter appears to be a slight advantage against the first (NLPCA as a technique to reduce the dimensionality), although the difference is not statistically significant. The two methods have in common that they use SOM for the

Table II Correlation coefficients between each region

Number of Cluster	PCA+HCA				PCA+SOM				NLPCA+HCA				NLPCA+SOM			
	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4
C1	1.00	0.18			1.00	0.18			1.00	0.20			1.00	0.18		
C2	0.18	1.00			0.18	1.00			0.20	1.00			0.18	1.00		
C1	1.00	0.15	0.23		1.00	0.18	0.69		1.00	-0.23	-0.77		1.00	0.18	0.69	
C2	0.23	1.00	0.90		0.18	1.00	0.16		-0.23	1.00	0.18		0.18	1.00	0.16	
C3	0.15	0.90	1.00		0.69	0.16	1.00		-0.77	0.18	1.00		0.69	0.16	1.00	
C1	1.00	0.20	0.16	0.69	1.00	0.17	0.16	-0.47	1.00	-0.23	0.60	0.69	1.00	0.18	-0.60	0.67
C2	0.20	1.00	0.90	0.16	0.17	1.00	-0.49	-0.52	-0.23	1.00	-0.17	-0.15	0.18	1.00	-0.15	0.16
C3	0.16	0.90	1.00	0.13	0.16	-0.49	1.00	0.69	0.60	-0.17	1.00	0.74	-0.60	-0.15	1.00	-0.73
C4	0.69	0.16	0.13	1.00	-0.47	-0.52	0.69	1.00	0.69	-0.15	0.74	1.00	0.67	0.16	-0.73	1.00

Table III Classification accuracy (%) of gauge stations – Discriminant analysis

Method	Number of Clusters		
	2	3	4
PCA+HCA	97.73	90.91	93.18
PCA+SOM	97.73	100.00	100.00
NLPCA+HCA	100.00	100.00	100.00
NLPCA+SOM	100.00	100.00	100.00

regionalization phase, a result expected given that SOM captures the cluster's shape as a non-linear method with high performance in cluster analysis on rainfall (Lin & Chen 2006, Farsadnia et al. 2014, Rau et al. 2017, Mannan et al. 2018, Markonis & Strnad 2019). Therefore, this study showed that SOM+NLPCA constitutes a powerful approach in the rainfall's regionalization process in an area of the tropical region with complex topographic conditions due to the Andes Mountain range and the proximity of the tropical Pacific Ocean.

Furthermore, according to the performance metrics, mainly CC, the adequate region's number is two, because the CC between C1 and C2 is low (0.18), indicating that they are heterogeneous (Table II). This result shows adequate geographical coherence in the spatial distribution of the 44 gauge stations in the study area (See Figure 6).

Region's characterization and rainfall analysis

The method NLPCA+SOM allowed establishing that Nariño has two homogeneous climatic: Andean Region (AR) and Pacific Region (PR) (See Figure 6). The spatial distribution of gauge stations showed that thirty-three gauge stations are over the AR, and eleven gauge stations are over the Pacific coast of Nariño PR in the southern Colombian Biogeographic Choco (Figure 7). The results of the regionalization process through the non-linear approaches are consistent with the 17 regions from mean monthly rainfall identified in Colombia by Guzmán et al. (2014) using monthly rainfall dataset of 408 gauge stations of IDEAM and PCA. It showed that Nariño is part of two climatic zones: region eight, which includes the south of Colombian Pacific and region ten that contain the Andean area of Nariño. Furthermore, the results also are consistent with the regionalization of monthly rainfall dataset of 1703 gauge stations managed by IDEAM in Colombia performed by Estupiñan (2016) through K-means, in which regions 1(16) matches with the AR(PR).

The two regions identified have different interannual rainfall variability patterns, which we document in this section. AR extends over eastern Nariño and includes the mountainous area where the average monthly rainfall is around 130 $mm.month^{-1}$. This region shows a bimodal annual cycle with wet seasons from March to May (MAM) and September to November (SON), and dry seasons from December to February (DJF) and from JJA (See Figure 7a). Furthermore, the second wet (dry) season SON(DJF) is more intense than the first season MAM(JJA). According to Guzmán et al. (2014), Poveda et al. (2011), and Schneider et al. (2014) the main factor that determines the bimodal regime of rainfall is the variation of moisture by the trade winds and the double passage of the ITCZ during the year, associated with the meridional migration of solar radiation. Moreover, other atmospheric mechanisms explain the rainfall cycle over the Andean areas, such as the westerlies winds, the complex Andean orography, local circulations (Espinoza et al. 2020), and the Easterly Waves that, according to Dominguez et al. (2020) to produce up to 50% of seasonal rainfall over northern South America.

The annual regime found is coherent with the analysis of seasonality of rainfall developed by Guzmán et al. (2014) and Estupiñan (2016).

Table IV. Results of Kolmogorov-Smirnov Test.

Number of Cluster	PCA+HCA				PCA+SOM				NLPCA+HCA				NLPCA+SOM			
	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4
C1	-	Ha	-	-	-	Ha	-	-	-	Ha	-	-	-	Ha	-	-
C2	Ha	-	-	-	Ha	-	-	-	Ha	-	-	-	Ha	-	-	-
C1	-	Ha	Ha	-	-	Ha	Ha	-	-	Ha	Ha	-	-	Ha	Ha	-
C2	Ha	-	Ha	-	Ha	-	Ha	-	Ha	-	Ha	-	Ha	-	Ha	-
C3	Ha	Ha	-	-	Ha	Ha	-	-	Ha	Ha	-	-	Ha	Ha	-	-
C1	-	Ha	Ha	Ha	-	Ha	Ha	Ha	-	Ha	Ho	Ha	-	Ha	Ha	Ha
C2	Ha	-	Ha	Ha	Ha	-	Ha	Ha	Ha	-	Ha	Ha	Ha	-	Ha	Ha
C3	Ha	Ha	-	Ha	Ha	Ha	-	Ha	Ho	Ha	-	Ha	Ha	Ha	-	Ha
C4	Ha	Ha	Ha	-	Ha	Ha	Ha	-	Ha	Ha	Ha	-	Ha	Ha	Ha	-

Ho: Distributions are the same;
Ha: Distributions are different.

Table V Ranks of clustering methods found for performance metric.

Performance metrics	PCA+HCA	PCA+SOM	NLPCA+HCA	NLPCA+SOM
CC_2	2	2	4	2
CC_3	4	2.5	1	2.5
CC_4	3	2	4	1
DA_2	3.5	3.5	1.5	1.5
DA_3	4	2	2	2
DA_4	4	2	2	2
$K\text{-}S_2$	2.5	2.5	2.5	2.5
$K\text{-}S_3$	2.5	2.5	2.5	2.5
$K\text{-}S_4$	2	2	4	2
Total Rank	27.5	21	23.5	18
R	3.06	2.33	2.61	**2.00**

They founded and described bimodal regimen in the Andean region of Nariño, and is consistent with the analysis performed recently by Urrea et al. (2019) in which explains that in Colombia the beginning of the first rainy season starts in the south of the Andean region in February and moves north in April. In contrast, the beginning of the second wet season occurs when the ITCZ moves to the south, this season starts in August at the north, but just in October, the season begins at the south of the region. Regarding the intensity of dry and wet periods, Urrea et al. (2019) reported that in the bimodal cycle of Colombia's Andean region, the second wet season is more intense than the first, while the dry season varies according to on location. These results also are consistent with Poveda & Mesa (2000) and Yepes et al. (2019). They reported that during SON the CJ modulates the wet season over western and central Colombia, given that the CJ

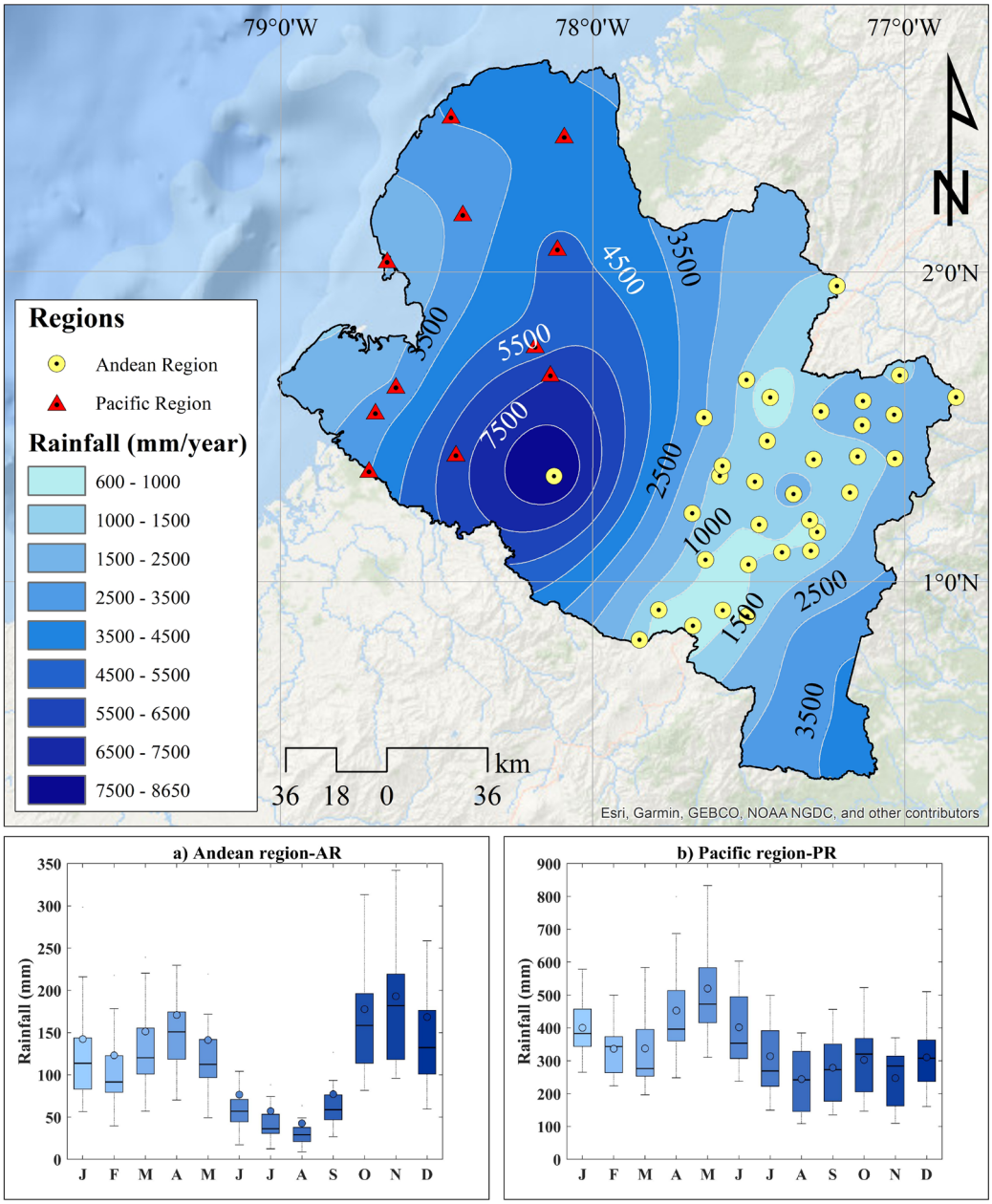

Figure 7. Annual means rainfall in Nariño and monthly rainfall regime (1983-2016) for the two identified regions. a). Andean region and b) Pacific region. Lower and upper box boundaries are 25th and 75th percentiles, respectively, line inside the box is median, lower and upper error lines are 10th and 90th percentiles, respectively, the hollow point is the mean.

exhibits an annual cycle with strengthening in SON and weakening in MAM. According to Poveda et al. (2014), the intensity of the wet (dry) seasons can increase during wet (dry) years linked to La Niña (El Niño) event of ENSO, and according to Dominguez et al. (2020), under ENSO neutral conditions the Easterly Waves contribution to interannual variability can become important in the north of South America.

Otherwise, PR covers the coastal area of Nariño, where the average monthly rainfall is around 350 mm. month^{-1}, which is twice the mean monthly rainfall of AR. This zone exhibits a unimodal annual cycle, that is, has one rainy from April to July, and one dry season from August to March (See Figure 7b). This result is consistent with the findings of Guzmán et al. (2014), Urrea et al. (2019), and Dominguez et al. (2020). They

described that the rainy season in Colombia occurs at the beginning of the year in the southernmost in early November and moves to the north, where it starts in May, according to the ITCZ's and the Easterly Waves movement. Besides, the second half of the year in this region is less rainy, due to the ITCZ's northernmost location over the continent and the eastern equatorial Pacific. The dry season starts in the South of Colombia May and beginning June and moves to the north where starts ending November and beginning December (Poveda et al. 2011, Córdoba-Machado et al. 2015, Urrea et al. 2019).

All the above results are consistent with the analysis seasonality of rainfall in Colombia developed by Urrea et al. (2019). They found that in Colombia's Pacific region occurs mainly the unimodal regime, which is characteristic in areas with low elevations. Another feature of this region is that the rainfall's intensities are higher than recorded in the AR. That behavior linked to low-level jets brings much moisture to this region as the CJ and the Caribbean low-level Jet that crosses the Isthmus of Panama by recurving and penetrating the Pacific Coast of Colombia (Amador et al. 2006, Poveda & Mesa 2000).

Finally, Figure 7 depicts the spatial distribution of annual mean rainfall, which indicates one core of intense rainfall from 7000 to 8500 mm. year^{-1}. These findings are consistent with Poveda et al. (2004) which used monthly-multiannual precipitation dataset from 411 meteorological stations in the Colombian Pacific region. They established ecogeographic sub-regions in the Choco Biogeographic Colombian, among which the South Pacific sub-region in the lower parts of the basins of the Patía and Mira rivers.

Furthermore, it is coherent with Cerón et al. (2021), who found three central nuclei of higher rainfall in the Choco Biogeographic Colombian through different methods of interpolation. The lowest intensity nucleus (from 3000 to 7000 mm. year^{-1}) coincides with the core shown in Figure 7. According to Poveda & Mesa (1999), there is a strong correlation between the rainfall of this region with the interaction between the surface winds of the Pacific Ocean and the eastern trade winds on the western cordillera. These come loaded with moisture, interact with the trade winds of the Andes, which, combined with the effect of surface warming and orographic rise, produce a highly unstable profile when colliding with the windward of the Andes. This region, is one of the rainiest places on Earth, the record rainfall is explained by the CJ, enhanced by atmosphere-ocean-land surface interactions (Poveda & Mesa 1999, Espinoza et al. 2020).

The described conditions above favor deep convection, the rise of humid air, the high amounts of condensation, which lead to an increase in the rainfall over the foothills of the Pacific plain (See Figure 7 and Figure 8). On the other hand, the rainfall regime on the eastern slope of the Andes Mountain Range is strongly influenced by moisture from the Amazon ecosystem, characterized by abundant water supply and a continuous process of evapotranspiration of the vegetation cover of this area with the overlying atmosphere (Barreiro & Díaz 2011). In the AR, 600 mm annual rainfall occurs, this is mainly due to the Foehn effect defined as strong, warm, and very dry winds descending in the leeside of mountains (Brinkmann 1971). This effect is induced by orography, where the warm air mass ascent on the windward slopes so the air cools, leading to condensation and latent heat release, allowing the formation of clouds and orographic rains over the windward slope; the rainfall removes the condensed water so that descent on the leeside is dry, which increases the warming, lead to higher leeside temperatures and the consequent diminishing of the moisture (Elvidge & Renfrew 2016).

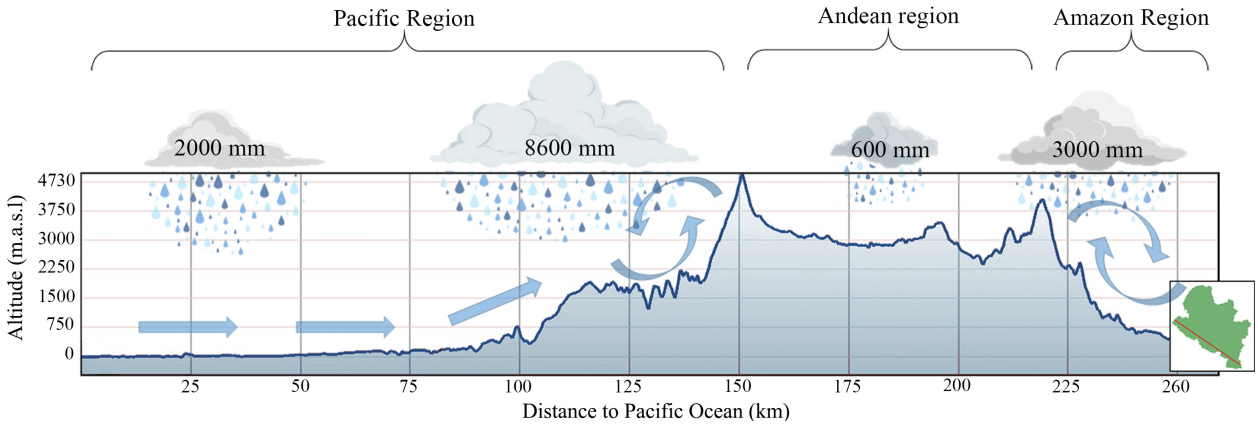

Figure 8. Cross-section profiles of Nariño showing the influence of the Tropical Pacific Ocean, Andes mountain, and Amazon region.

CONCLUSIONS

Nariño's location in the Intertropical Convergence Zone (ITCZ), near to the Pacific Ocean and with the presence of the Andes Mountain, makes it challenging to understand spatially and temporally the variability of rainfall and other climatic parameters which are relevant for the water management. Therefore, considering the previously mentioned, four methods for regionalization of monthly rainfall of 34 years in Nariño (Southwestern Colombia) were tested and compared. The two best techniques have proved to be Non-Linear Principal Component Analysis and Self-Organizing Maps (NLPCA+SOM) and Principal Component Analysis-PCA+SOM. However, the ability of the non-linear approach to capture non-linear relationships makes them the best option. Three performance metrics established the accuracy of the regionalization process: the Pearson's correlation coefficient (CC) of the first PCs between each region, the Discriminant Analysis (DA), and the Kolmogorov-Smirnov Test (K-S). The best technique, considering each metric, was chosen by applying the Wilcoxon Test.

The clustering of stations rainfall by NLPCA+SOM led to identifying two homogeneous regions with different seasonal rainfall patterns. The identified clusters correspond to two: Andean Region (AR) and Pacific Region (PR) with a bimodal and unimodal annual cycle, respectively. Both groups showed that the rainfall pattern is mainly influenced by the ITCZ's movement, the complex region's orography, the Choco low level jet, the Easterly Waves, and the ENSO phenomenon. The analysis of the annual rainfall cycle of AR indicated a reduction in two-quarters: From December to February and from June to August; it also shows an increase between March to May and from September to November. Meanwhile, the PR analysis showed one wet season from April to July and a one-dry season from August to March.

The intensity of rainfall in PR was higher than recorded in the AR due to the Pacific Ocean's surface winds interacting with the eastern trade winds on the western Andes cordillera. These interactions with the surface warming effect, and the orographic rise conducive to the deep convection, the elevation of moisture, the high amounts of condensation, therefore, the high rainfall. The obtained results are relevant and useful for accurate characterization of the basin's hydrology and better water and agricultural planning in the region.

Future works will be oriented into the study of rainfall variability in each region identified here and their teleconnections with large-scale climate indices, e.g., detecting relationships to ENSO, and in this way, allow the construction of better models for water management.

Acknowledgments

The authors thank to Universidad del Valle for supporting the research project CI 21010, and Colciencias for funding the research project "Análisis de eventos extremos de precipitación asociados a variabilidad y cambio climático para la implementación de estrategias de adaptación en sistemas productivos agrícolas de Nariño". The first author received a scholarship from program Fortalecimiento de Capacidades Regionales en Investigación, Desarrollo Tecnológico e Innovación de Nariño. Colciencias supported the second author, and Universidad del Valle helped the fifth author. Furthermore, the authors thank the IREHISA and PSI research groups for the support received during the development of this research paper. Finally, the authors express their thanks to IDEAM for providing the database containing the monthly rainfall in the Department of Nariño.

REFERENCES

AGARWAL A, MAHESWARAN R, KURTHS J & KHOSA R. 2016. Wavelet Spectrum and self-organizing maps-based approach for hydrologic regionalization-a case study in the western United States. Water Res Manag 30: 4399-4413.

AMADOR J, ALFARO E, LIZANO O & MAGAÑA V. 2006. Atmospheric forcing of the eastern tropical Pacific: A review. Prog Ocean 69: 101-142.

ARIAS P, MARTÍNEZ J & VIEIRA S. 2015. Moisture sources to the 2010–2012 anomalous wet season in northern South America. Climate Dynamics 45: 2861-2884.

ASONG Z, KHALIQ M & WHEATER H. 2015. Regionalization of precipitation characteristics in the Canadian Prairie Provinces using large-scale atmospheric covariates and geophysical attributes. Stoch Environ Res Risk Assess 29: 875-892.

BARREIRO M & DÍAZ N. 2011. Land–atmosphere coupling in El Niño influence over South America. Atmospheric Science Letters 12: 351-355.

BEDOYA-SOTO J, ARISTIZABAL E, CARMONA A & POVEDA G. 2019. Seasonal Shift of the Diurnal Cycle of Rainfall Over Medellin's Valley, Central Andes of Colombia (1998-2005). Front Earth Sci 7: 92.

BRINKMANN W. 1971. What is a foehn? Weather 26: 230-240.

CANCHALA T, ALFONSO-MORALES W, CERÓN W, CARVAJAL-ESCOBAR Y & CAICEDO-BRAVO E. 2020. Teleconnections between monthly rainfall variability and large-scale climate indices in Southwestern Colombia. Water 12: 1863.

CANCHALA T, CARVAJAL-ESCOBAR Y, ALFONSO-MORALES W, CERÓN W & CAICEDO-BRAVO E. 2019. Estimation of missing data of monthly rainfall in southwestern Colombia using artificial neural networks. Data in Brief 26: 104517.

CARMONA A & POVEDA G. 2012. Aplicación de la transformada de Hilbert-Huang en la detección de modos de variabilidad hidroclimática en Colombia. Dyna, 79.

CARVAJAL Y & SEGURA J. 2004. Análisis de variabilidad de datos medioambientales aplicando funciones ortogonales empíricas o componentes principales. Ingeniería de Recursos Naturales y del Ambiente 1: 4-11.

CERÓN W, ANDREOLI R, KAYANO M, FERREIRA DE SOUZA R, CANCHALA T & CARVAJAL Y. 2021. Comparison of spatial interpolation methods for annual and seasonal rainfall in two hotspots of biodiversity in South America. An Acad Bras Cienc 93: e20190674. https://doi.org/10.1590/0001-3765202120190674

CERÓN W, ANDREOLI R, KAYANO M, FERREIRA DE SOUZA R, JONES C & CARVALHO L. 2020. The Influence of the Atlantic Multidecadal Oscillation on the Choco Low-Level Jet and Precipitation in Colombia. Atmosphere 11: 174.

CÓRDOBA-MACHADO S, PALOMINO-LEMUS R, GÁMIZ-FORTIS S, CASTRO-DÍEZ Y & ESTEBAN-PARRA M. 2015. Influence of tropical Pacific SST on seasonal precipitation in Colombia: prediction using El Niño and El Niño Modoki. Climate Dynamics 44: 1293-1310.

CHEN L & HONG Y. 2012. Regional Taiwan rainfall frequency analysis using principal component analysis, self-organizing maps and L-moments. Hydrol Res 43: 275-285.

CHEN L, LIN G & HSU C. 2011. Development of design hyetographs for ungauged sites using an approach combining PCA, SOM and kriging methods. Water Res Manag 25: 1995-2013.

DARAND M & DANESHVAR M. 2014. Regionalization of precipitation regimes in Iran using principal component analysis and hierarchical clustering analysis. Environ Proc 1: 517-532.

DEMŠAR J. 2006. Statistical comparisons of classifiers over multiple data sets. J Machine Learn Res 7: 1-30.

DIAZ L & MORALES M. 2002. Estadística multivariada: inferencia y métodos, Universidad Nacional de Colombia.

DOMINGUEZ C, DONE J & BRUYÈRE C. 2020. Easterly wave contributions to seasonal rainfall over the tropical Americas in observations and a regional climate model. Climate Dynamics 54: 191-209.

ELVIDGE A & RENFREW I. 2016. The causes of foehn warming in the lee of mountains. Bull Am Meteorol Soc 97: 455-466.

ESPINOZA J, GARREAUD R, POVEDA G, ARIAS P, MOLINA-CARPIO J, MASIOKAS M, VIALE M & SCAFF L. 2020. Hydroclimate of the Andes Part I: Main climatic features. Front Earth Sci 8: 64.

ESTUPIÑAN A. 2016. Estudio de la variabilidad espacio temporal de la precipitación en Colombia. PhD dissertation. Universidad Nacional de Colombia: http://bdigital.unal.edu.co/54014/1/1110490004.2016.pdf (last access: November 26, 2019).

FARSADNIA F, KAMROOD M, NIA A, MODARRES R, BRAY M, HAN D & SADATINEJAD J. 2014. Identification of homogeneous regions for regionalization of watersheds by two-level self-organizing feature maps. J Hydrol 509: 387-397.

FAZEL N, BERNDTSSON R, UVO C, MADANI K & KLØVE B. 2018. Regionalization of precipitation characteristics in Iran's Lake Urmia basin. Theor Appl Climatol 132: 363-373.

FRALEY C & RAFTERY A. 2002. Model-based clustering, discriminant analysis, and density estimation. J Am Stat Assoc 97: 611-631.

GARREAUD R, VUILLE M, COMPAGNUCCI R & MARENGO J. 2009. Present-day south american climate. Palaeogeography, Palaeoclimatology, Palaeoecology 281: 180-195.

GOCIC M & TRAJKOVIC S. 2014. Spatio-temporal patterns of precipitation in Serbia. Theor Appl Climatol117: 419-431.

GONG X & RICHMAN M. 1995. On the application of cluster analysis to growing season precipitation data in North America east of the Rockies. J Climate 8: 897-931.

GUERRERO E, DE KEIZER O & CÓRDOBA R. 2006. La aplicación del enfoque ecosistémico en la gestión de los recursos hídricos: un análisis de estudios de caso en América Latina. IUCN Ecuador 78.

GUZMÁN A. 2016. Análisis del comportamiento espacio temporal de la precipitación en la region pacifica colombiana en presencia de los fenomenos del El Niño y La Niña, en el periodo 1983-2009. Dissertation. Universidad de Nariño: http://sired.udenar.edu.co/1584/ (last access: May 20, 2019). (Unpublished).

GUZMÁN D, RUÍZ J & CADENA M. 2014. Regionalización de Colombia según la estacionalidad de la precipitación media mensual, a través análisis de componentes principales (acp). Grupo de Modelamiento de Tiempo, Clima y Escenarios de Cambio Climático: Bogotá: http://www.ideam.gov.co/documents/21021/21141/Regionalizacion+de+la+Precipitacion+Media+Mensual/1239c8b3-299d-4099-bf52-55a414557119 (last access: November 26, 2019).

HERSHEY R, MIZELL S & EARMAN S. 2010. Chemical and physical characteristics of springs discharging from regional flow systems of the carbonate-rock province of the Great Basin, western United States. Hydrogeol J 18: 1007-1026.

HERVADA-SALA C & JARAUTA-BRAGULAT E. 2004. A program to perform Ward's clustering method on several regionalized variables. Comput Geosci 30: 881-886.

HOYOS I, DOMINGUEZ F, CAÑÓN-BARRIGA J, MARTÍNEZ J, NIETO R, GIMENO L & DIRMEYER P. 2018. Moisture origin and transport processes in Colombia, northern South America. Climate Dynamics 50: 971-990.

HSIEH W. 2001. Nonlinear principal component analysis by neural networks. Tellus A: Dynamic Meteorology and Oceanography 53: 599-615.

HSU K & LI S. 2010. Clustering spatial–temporal precipitation data using wavelet transform and self-organizing map neural network. Adv Water Res 33: 190-200.

JARAMILLO L, POVEDA G & MEJÍA J. 2017. Mesoscale convective systems and other precipitation features over the tropical Americas and surrounding seas as seen by TRMM. International J Climatol 37: 380-397.

JOLLIFFE I. 1986. Principal Component Analysis, Springer, New York 271.

JOLLIFFE I & CADIMA J. 2016. Principal component analysis: a review and recent developments. Phil Trans R Soc A 374: 20150202.

KOHONEN T. 1982. Self-organized formation of topologically correct feature maps. Biol Cybern 43: 59-69.

KOHONEN T. 2001. Self-Organizing Maps. Springer Series in Information Sciences 30: 502.

KRAMER M. 1991. Nonlinear principal component analysis using autoassociative neural networks. AIChE Journal 37: 233-243.

LI-JUAN C, DE-LIANG C, HUI-JUN W & JING-HUI Y. 2009. Regionalization of precipitation regimes in China. Atmospher Ocean Sci Lett 2: 301-307.

LIN G & CHEN L. 2006. Identification of homogeneous regions for regional frequency analysis using the self-organizing map. J Hydrol 324: 1-9.

MAGALLANES Q, VALDEZ C, MÉNDEZ G, MORENO B, MEDINA G & BLANCO M. 2015. Fractal analysis of monthly evaporation and precipitation time series at central Mexico. Terra Latinoamericana 33: 221-231.

MANNAN A, CHAUDHARY S, DHANYA C & SWAMY A. 2018. Regionalization of rainfall characteristics in India incorporating climatic variables and using self-organizing maps. ISH Journal of Hydraulic Engineering 24: 147-156.

MARKONIS Y & STRNAD F. 2019. Representation of European hydroclimatic patterns with self-organizing maps. The Holocene 8:1155-1162.

MARTINS D, RAZIEI T, PAULO A & PEREIRA L. 2012. Spatial and temporal variability of precipitation and drought in Portugal. Natural Hazards & Earth System Sciences 12:1493-1501.

MIRÓ J, CASELLES V & ESTRELA M. 2017. Multiple imputation of rainfall missing data in the Iberian Mediterranean context. Atmospheric research 197: 313-330.

NOURANI V, BAGHANAM A, VOUSOUGHI F & ALAMI M. 2012. Classification of groundwater level data using SOM to develop ANN-based forecasting model. Int J Soft Comput Eng 2: 2231-2307.

POVEDA G. 2004. La hidroclimatología de Colombia: una síntesis desde la escala inter-decadal hasta la escala diurna. Rev Acad Colomb Cienc 28: 201-222.

POVEDA G, ÁLVAREZ D & RUEDA Ó. 2011. Hydro-climatic variability over the Andes of Colombia associated with ENSO: a review of climatic processes and their impact on one of the Earth's most important biodiversity hotspots. Climate Dynamics 36: 2233-2249.

POVEDA G, ESPINOZA J, ZULUAGA M, SOLMAN S & GARREAUD R. 2020. High Impact Weather Events in the Andes. Front Earth Sci 8: 162.

POVEDA G, JARAMILLO L & VALLEJO L. 2014. Seasonal precipitation patterns along pathways of South American low-level jets and aerial rivers. Water Resour Res 50: 98-118.

POVEDA G & MESA O. 1997. Feedbacks between hydrological processes in tropical South America and large-scale ocean–atmospheric phenomena. J Climate 10: 2690-2702.

POVEDA G & MESA O. 1999. La corriente de chorro superficial del Oeste ("del Chocó") y otras dos corrientes de chorro en Colombia: climatología y variabilidad durante las fases del ENSO". Rev Acad Colomb Cienc 23: 517-528.

POVEDA G & MESA O. 2000. On the existence of Lloró (the rainiest locality on Earth): Enhanced ocean-d-atmosphere interaction by a low-level jet. Geophys Res Lett 27: 1675-1678.

POVEDA G, WAYLEN P & PULWARTY R. 2006. Annual and inter-annual variability of the present climate in northern South America and southern Mesoamerica. Palaeogeogr Palaeoclimatol Palaeoecol 234: 3-27.

POVEDA I, ROJAS C, RUDAS A & RANGEL O. 2004. El Chocó biogeográfico: ambiente físico. Colombia diversidad biótica IV, El Chocó biogeográfico/Costa Pacífica: 1-21.

PREISENDORFER R. 1988. Principal component analysis in meteorology and oceanography. Elsevier Sci Publ 17: 425.

PUERTAS OL & CARVAJAL Y. 2008. Incidence of El Niño southern oscillation in the precipitation and the temperature of the air in Colombia, using Climate Explorer. Ingeniería y Desarrollo: 104-118.

RAMI M, DE CAMPOS V & FERREIRA N. 2005. Artificial neural network technique for rainfall forecasting applied to the Sao Paulo region. J Hydrol 301: 146-162.

RAU P, BOURREL L, LABAT D, MELO P, DEWITTE B, FRAPPART F, LAVADO W & FELIPE O. 2017. Regionalization of rainfall over the Peruvian Pacific slope and coast. Int J Climatol 37: 143-158.

RAZIEI T. 2018. A precipitation regionalization and regime for Iran based on multivariate analysis. Theor Appl Climatol 131: 1429-1448.

RAZIEI T, BORDI I & PEREIRA L. 2008. A precipitation-based regionalization for Western Iran and regional drought variability. Hydrol Earth System Sci 12: 1309-1321.

SAMUEL J, COULIBALY P & METCALFE R. 2011. Estimation of continuous streamflow in Ontario ungauged basins: comparison of regionalization methods. J Hydrol Eng 16: 447-459.

SANTOS E, LUCIO P & SILVA C. 2015. Precipitation regionalization of the Brazilian Amazon. Atmosph Sci Lett 16: 185-192.

SATYANARAYANA P & SRINIVAS V. 2008. Regional frequency analysis of precipitation using large-scale atmospheric variables. J Geophys Res Atmospheres: 113.

SATYANARAYANA P & SRINIVAS V. 2011. Regionalization of precipitation in data sparse areas using large scale atmospheric variables–A fuzzy clustering approach. J Hydrol 405: 462-473.

SCHNEIDER T, BISCHOFF T & HAUG G. 2014. Migrations and dynamics of the intertropical convergence zone. Nature 513: 45.

SCHOLZ M, FRAUNHOLZ M & SELBIG J. 2008. Nonlinear principal component analysis: neural network models and applications. Principal manifolds for data visualization and dimension reduction. Springer.

SCHOLZ M, KAPLAN F, GUY C, KOPKA J & SELBIG J. 2005. Non-linear PCA: a missing data approach. Bioinformatics 21: 3887-3895.

SCHOLZ M & VIGÁRIO R. 2002. Nonlinear PCA: a new hierarchical approach. ESANN: 439-444.

SEO S, BHOWMIK R, SANKARASUBRAMANIAN A, MAHINTHAKUMAR G & KUMAR M. 2019. The role of cross-correlation between precipitation and temperature in basin-scale simulations of hydrologic variables. J Hydrol 570: 304-314.

SERNA L, ARIAS P & VIEIRA S. 2018. Las corrientes superficiales de chorro del Chocó y el Caribe durante los eventos de El Niño y El Niño Modoki. Rev Acad Colomb Cienc Exact Fís Natural 42: 410-421.

SHEN C. 2019. The influence of a scaling exponent on ρDCCA: A spatial cross-correlation pattern of precipitation records over eastern China. Physica A: Statistical Mechanics and its Applications 516: 579-590.

SHEN S, WIED O, WEITHMANN A, REGELE T, BAILEY B & LAWRIMORE J. 2016. Six temperature and precipitation regimes of the contiguous United States between 1895 and 2010: a statistical inference study. Theor Appl Climatol 125: 197-211.

SHESKIN D. 2003. Handbook of parametric and nonparametric statistical procedures, Chapman and Hall/CRC.

SRINIVAS V 2013. Regionalization of precipitation in India–a review. J Indian Inst Sci 93: 153-162.

TOŠIĆ I, ZORN M, ORTAR J, UNKAŠEVIĆM, GAVRILOV M & MARKOVIĆ S. 2016. Annual and seasonal variability of precipitation and temperatures in Slovenia from 1961 to 2011. Atmosph Res 168: 220-233.

TOTH E. 2013. Catchment classification based on characterisation of streamflow and precipitation time series. Hydrol Earth Syst Sci 17:1149-1159.

URREA V, OCHOA A & MESA O. 2019. Seasonality of Rainfall in Colombia. Water Resour Res 55: 4149-4162.

WANG N & YIN J. 2019. Self-organizing map network-based precipitation regionalization for the Tibetan Plateau and regional precipitation variability. Theor Appl Climatol135: 29-44.

WANG Y, HUYNH G & WILLIAMSON C. 2013. Integration of Google Maps/Earth with microscale meteorology models and data visualization. Comput Geosci 61: 23-31.

WILCOXON F. 1992. Individual comparisons by ranking methods. Breakthroughs in statistics. Springer, New York.

WON C, LIEW J, YUSOP Z, ISMAIL T, VENNEKER R & UHLENBROOK S. 2016. Rainfall characteristics and regionalization in Peninsular Malaysia based on a high resolution gridded data set. Water 8: 500.

YEPES J, POVEDA G, MEJÍA J, MORENO L & RUEDA C. 2019. CHOCO-JEX: a research experiment focused on the CHOCO low-level jet over the far Eastern Pacific and Western Colombia. Bull Am Meteorol Soci 100: 779-796.

How to cite

CANCHALA T, OCAMPO-MARULANDA C, ALFONSO-MORALES W, CARVAJAL-ESCOBAR Y, CERÓN WL & CAICEDO-BRAVO E. 2022. Techniques for monthly rainfall regionalization in Southwestern Colombia. An Acad Bras Cienc 94: e20201000. DOI 10.1590/0001-3765202220201000.

Manuscript received on June 30, 2020;
accepted for publication on March 26, 2021

TERESITA CANCHALA[1]
https://orcid.org/0000-0002-5208-5515

CAMILO OCAMPO-MARULANDA[1]
https://orcid.org/0000-0002-3813-8780

WILFREDO ALFONSO-MORALES[2]
https://orcid.org/0000-0002-3091-6082

YESID CARVAJAL-ESCOBAR[1]
https://orcid.org/0000-0002-2014-4226

WILMAR L. CERÓN[3]
https://orcid.org/0000-0003-1901-9572

EDUARDO CAICEDO-BRAVO[2]
https://orcid.org/0000-0003-0727-2917

[1]Universidad del Valle, Grupo de Investigación en Ingeniería de los Recursos Hídricos y Suelos (IREHISA), Escuela de Ingeniería de los Recursos Naturales y del Ambiente, Calle 13, No. 100-00, Campus Meléndez, Cali, Colombia

[2]Universidad del Valle, Grupo de Investigación Percepción y Sistemas Inteligentes (PSI), Escuela de Ingeniería Eléctrica y Electrónica, Calle 13, No. 100-00, Campus Meléndez, Cali, Colombia

[3]Universidad del Valle, Departamento de Geografía, Facultad de Humanidades, Calle 13, No. 100-00, Campus Meléndez, Cali, Colombia

Correspondence to: **Teresita Canchala Nastar**
E-mail: teresita.canchala@correounivalle.edu.co

Author Contributions

Conceptualization—T.C., C.O.-M., W.A.-M., W.L.C., and Y.C.-E; methodology—T.C., C.O.-M., Y.C.-E., W.A.-M., and W.L.C; validation—T.C., C.O.-M., W.A.-M., and Y.C.-E; formal analysis— T.C., W.A.-M., W.L.C., and Y.C.-E; investigation— T.C., C.O.-M., W.A.-M., W.L.C., and Y.C.-E; data curation—T.C, and W.A.-M; original draft preparation—T.C., C.O.-M., W.A.-M., W.L.C., and Y.C.-E; reviewing and editing—T.C., W.A.-M., W.L.C., Y.C.-E., and E.C.-B; visualization—T.C., C.O.-M., W.A.-M., W.L.C., Y.C.-E., and E.C.-B; supervision—W.A.-M., and Y.C.-E. All authors have read and agreed to the published version of the manuscript.