

Identification of robust catchment classification methods for Sahelian watersheds

Pedram Darbandsari^{a,*}, Paulin Coulibaly^{a,b,c}, Jafet C.M. Andersson^d

^a School of Earth, Environment and Society, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4L7, Canada

^b Department of Civil Engineering, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4L7, Canada

^c United Nations University Institute for Water, Environment, and Health, Hamilton, ON L8P 0A1, Canada

^d Hydrological Research and Development Unit, Swedish Meteorological and Hydrological Institute (SMHI), Folkborgsvägen 17, 60176, Sweden

ARTICLE INFO

Keywords:

Watershed classification
K-means algorithm
Self-Organizing Maps
Principal Component Analysis (PCA)
Non-linear Principal Component Analysis (NLPCA)
Sahelian watershed
West Africa
Bagnouls-Gausson (BGI) index

ABSTRACT

Study region: The study was conducted in a Sahelian watershed located in Burkina Faso (West Africa).

Study focus: In this study, an inter-comparison procedure is proposed to investigate the effects of implementing various sets of explanatory variables and clustering algorithms on developing hydrologically homogenous regions. Six different sets of explanatory variables considered in this framework are generated using the combinations of topographic, land-use, climatic, and hydrological attributes. Also, seven different linear and nonlinear clustering techniques are implemented using the combinations of Principal Component Analysis (PCA), Non-linear Principal Component Analysis (NLPCA), Self-Organizing Maps (SOM), and K-means algorithm. The mean and maximum annual runoff are considered as two variables of interest for conducting a comparison and identifying the most robust classification methods.

New hydrological insights for the region: The study results indicate that the monthly Bagnouls-Gausson index (BGI) is the most robust set of explanatory variables to be used for identifying the hydrologically homogenous regions considering both mean and maximum annual runoff. Additionally, compared with BGI, the combination of topographic and land-use attributes can provide competitive results while the land-use attributes alone cannot capture the hydrological heterogeneity of the catchments. Moreover, interestingly, the comparison results show that regardless of its simplicity, the K-means algorithm is superior to the other clustering techniques in terms of generating hydrologically homogenous regions based on the monthly BGI.

1. Introduction

Classification of catchments into hydrologically similar groups is a principal issue for operational hydrology and water resources management (Toth, 2013). The main goals of catchment classification are to find the patterns in the heterogeneity of water systems and classify the catchments into different groups with limited variability within each class (Jehn et al., 2020; Sharghi et al., 2018). Catchment classification is a valuable step in understanding water systems (Ley et al., 2011). Besides the development of generalization of hydrologic systems, one of the most advantageous applications of catchment classification is in regionalization studies (Razavi and Coulibaly, 2013; 2022). In many regionalization studies, the first step involves categorizing catchments into

* Corresponding author.

E-mail addresses: darbandp@mcmaster.ca (P. Darbandsari), couliba@mcmaster.ca (P. Coulibaly).

<https://doi.org/10.1016/j.ejrh.2024.102067>

Received 19 December 2023; Received in revised form 13 July 2024; Accepted 6 November 2024

Available online 23 November 2024

2214-5818/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

hydrologically homogeneous regions to facilitate the transfer of hydrological information from gauged to ungauged catchments (Di Prinzio et al., 2011; Mostofi Zadeh, 2012). Regional frequency analysis is one of the most common examples of regionalization studies that requires the identification of homogeneous regions. If the objective is flood frequency analysis, the variable of interest to be predicted is flood peaks, while in the case of water availability assessment, the mean or low flow percentiles would be the streamflow indices of interest (Sharghi et al., 2018; Toth, 2013).

In general, catchment classification studies are mostly based on either the watersheds' spatial, geospatial, and climatic properties (e.g. Ahani et al., 2020; Chérif and Bargaoui, 2013; Farsadnia et al., 2014; Razavi and Coulibaly, 2017) or their streamflow metrics as their hydrologic response behaviors (Abdolhay et al., 2012; Boscarello et al., 2016; Ercan et al., 2008; Sharghi et al., 2018), or both (Donnelly et al., 2016). As the proximity of watershed characteristics may suggest the similarity of hydrologic behavior of the catchment, they can be used for the purpose of regionalization (Ssegane et al., 2012). On the other hand, although streamflow metrics

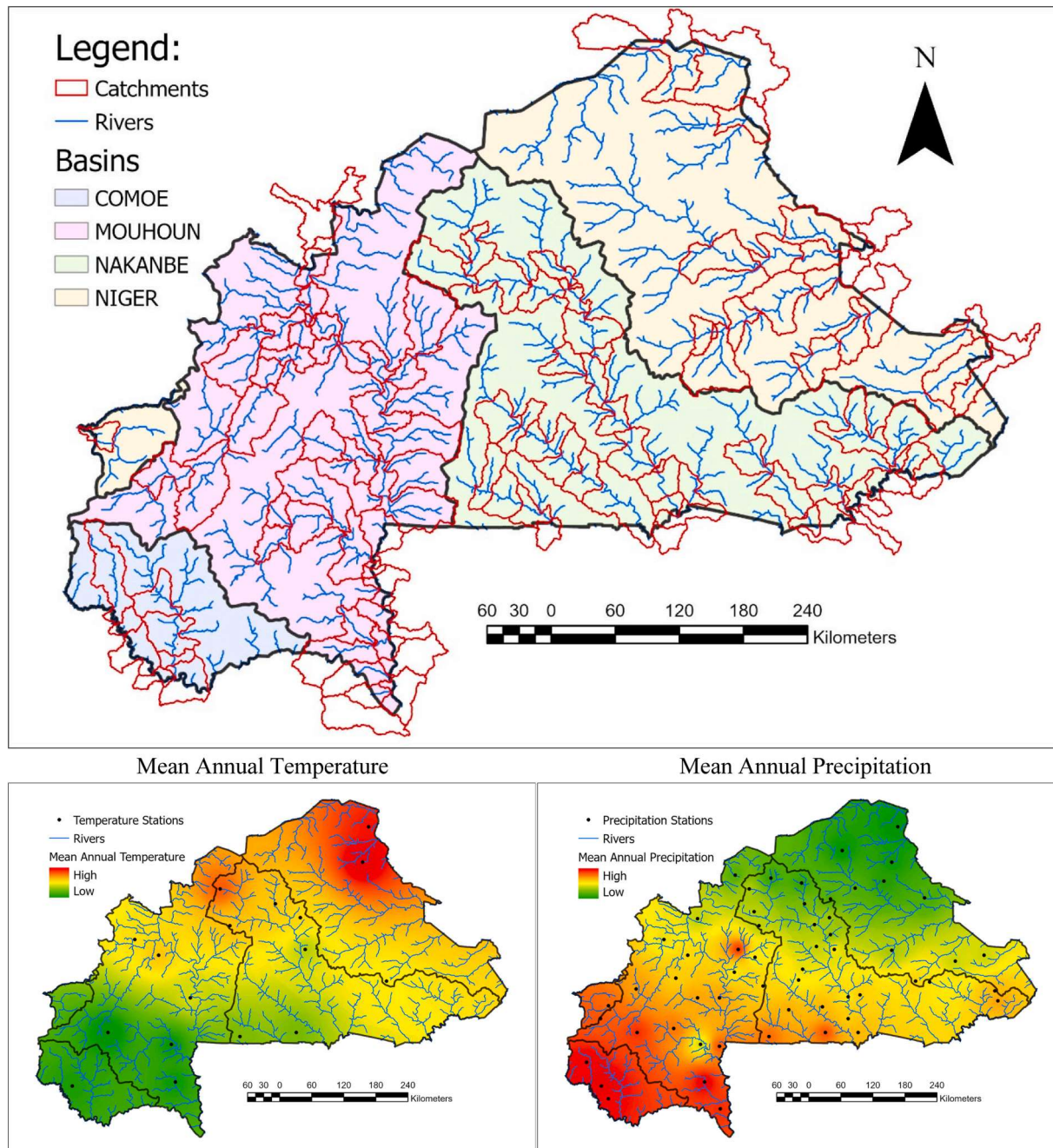


Fig. 1. Study area. Location of catchments and the variation of mean annual precipitation and temperature over the region.

can be reliable variables for estimating hydrological similarity, they are not proper for ungauged regions where limited streamflow data is available (Razavi and Coulibaly, 2013). Furthermore, the implementation of multivariate statistical analysis, such as principal component analysis (PCA), is proposed in order to reduce the high dimensionality of explanatory variables which may enhance the quality of the classification results (e.g. Di Prinzio et al., 2011; Farsadnia et al., 2014; Ssegane et al., 2012; Toth, 2013; Wolfe et al., 2019).

The most common procedure being used in the literature in the context of catchment classification is the clustering technique (Boscarello et al., 2016). Maximizing the similarity of catchments within each group while minimizing the similarity between two clusters is the main idea of cluster analysis, which is in line with the concept of catchment classification (Ley et al., 2011). There are various conventional clustering algorithms and the K-means clustering algorithm (MacQueen, 1967) is one of the most commonly used ones for catchment classification purposes (e.g. Burn and Boorman, 1993; Burn and Goel, 2000; Lin and Chen, 2006; Sharghi et al., 2018; Ssegane et al., 2012). However, non-linear clustering techniques such as Self-Organizing Map (SOM), an unsupervised artificial neural network, have been recently implemented for classifying catchments (e.g. Boscarello et al., 2016; Farsadnia et al., 2014; Herbst and Casper, 2008; Zhang et al., 2022). Unlike K-means, which is a linear clustering technique, the SOM method can capture the non-linear patterns in data and besides classification, it can also be used for reducing dimensionality (Razavi and Coulibaly, 2013; Sharghi et al., 2018).

Although various studies applied different techniques for identifying homogenous watersheds, there is still no universally agreed upon watershed classification system (Salami and Buehler, 2020; Wagener et al., 2007). Some prior studies aimed to compare various classification approaches or explanatory variables but none concurrently considered the effects of both aspects. Some studies focused on the examination of explanatory variable selection (e.g., Ssegane et al., 2012), while others dedicated their attention to comparing clustering algorithms (e.g., Farsadnia, 2014; Di Prinzio, 2011; Razavi and Coulibaly, 2013; Sharghi et al., 2018). This underscores a knowledge gap regarding the interaction between explanatory variables and clustering algorithms in watershed classification.

Furthermore, some clustering techniques require specifying the number of clusters prior to the cluster analysis. Exploring various numbers of clusters and selecting the most robust ones based on clustering performance evaluation criteria was implemented in classification studies (Sharghi et al., 2018; Razavi and Coulibaly, 2013). The performance of the classification approach is highly related to the number of clusters used in cluster analysis. Therefore, considering this effect on the catchment classification inter-comparison process is necessary for gaining more reliable conclusions, which has been overlooked in previous studies.

This paper tries to overcome the aforementioned limitations in previous catchment classification comparison studies, by proposing an inter-comparison procedure where the effects of explanatory variables, the classification approaches, and the number of clusters on the performance of various catchment classification results are taken into account. By focusing on 116 catchments in Burkina Faso, the proposed inter-comparison framework thoroughly compared the performance of six sets of explanatory variables, including topographic, land-use, climate, and streamflow attributes using seven different clustering approaches, developed by integration of PCA, NLPCA, SOM, and K-means, and a varying number of clusters to reduce their effects on the performance of the classification techniques. The results are expected to contribute toward the development of a more widely applicable classification system by proposing a systematic comparison framework and offering recommendations on the effectiveness of various classification techniques. Moreover, the proposed framework also provides fresh insight into the most dominant drivers of catchment behavior by comparing various explanatory variables. The structure of the paper is as follows: in the next section, the proposed inter-comparison framework and all its underlying methods and materials, including the study area, are elaborated. Section 3 discusses the implementation of the framework and inter-comparison results and finally, in Section 4, the summary and key conclusions are presented.

2. Material and methods

2.1. Study area and datasets

The Sahelian watersheds in West Africa are in general poorly gauged and undergoing dynamic changes due to climate extremes and effects of anthropogenic activities. This is a region where adequate catchment classification methods are much needed for

Table 1
The catchments' topographic and land-cover attributes.

Catchment Attributes	Description	Unit	Range
Topographic	Latitude	Degree	9.4–15
	Longitude	Degree	–5.2–2.2
	Accumulated Area	km ²	3211–100356
	Elevation	m	168–397
	Slope	Degree	0.1–1.5
Land-use	Bare	%	9.3–26
	Crops	%	1.4–27.3
	Grass	%	0 – 6.2
	Shrub	%	0–30.2
	Tree	%	10.8 – 44.6
	Urban	%	30.7–55.0
	Water	%	0–3.1
	Wetland	%	0–6.1

regionalization studies because of the scarcity of observed streamflow data. Therefore, to implement the proposed inter-comparison procedure, watersheds located/partly located in Burkina Faso in West Africa are chosen (Fig. 1). The study area of interest is located between latitude 9° N and 15° N, and longitude 5° W and 3° E. The mean annual precipitation in the study region ranges from 430 mm in the northeast to more than 1100 mm in the southeast, while, the variation of the annual mean temperature is not that significant, ranging from 26° C in southwestern regions to 30° C in northeastern ones. To apply the catchment classification methods, 116 catchments with river orders of two or more were selected over the study regions. These catchments and their corresponding outlets are in one of the four main river basins in the country, including Comoé, Mouhoun, Nakanbe, and Niger (Fig. 1).

Using the available data (i.e. Digital elevation and land-cover maps), the topographic and land-use attributes of all catchments are scrutinized and extracted to be used as the variables for classification. The ranges of these attributes, presented in Table 1, indicate distinct physiographic characteristics of the catchments and along with different climatic features and suggest that the region does not have homogenous hydrological behaviour across its catchments. From the available meteorological gauging stations, the daily mean areal precipitation and temperature in each catchment are interpolated using the inverse distance weighting approach (American Society of Civil Engineers, 1996). The daily streamflow data were obtained from the output of the World-Wide HYPE hydrologic model adapted for West Africa (Lindström et al., 2010; Arheimer et al., 2020; Berg et al., 2021) at the outlets of the selected catchments. A total of 37 years of data from 1979 to 2015 are used in this study. Both maximum annual and mean annual runoff are considered as the variables of interest to examine the capability of different classification procedures in generating homogenous regions. The variability of the maximum and mean annual runoff over the study area is illustrated in Fig. 2.

2.2. Catchment classification inter-comparison framework

The proposed framework introduces an intercomparison procedure that assesses the effects of various explanatory variables and classification approaches on the performance of catchment classification results (Fig. 1). The framework begins with applying the discordancy measure proposed by Hosking and Wallis (1993, 2005) to identify and exclude highly discordant catchments. In the second step, the selected non-discordant catchments are classified into different groups using seven different clustering techniques, which are developed as an integration of various pre-processing and clustering methods. This step also takes into account the effect of the explanatory variables and the pre-specified number of groups on the classification results by considering six different sets of variables and using various numbers of clusters, ranging from four to ten. Given six sets of explanatory variables, seven different

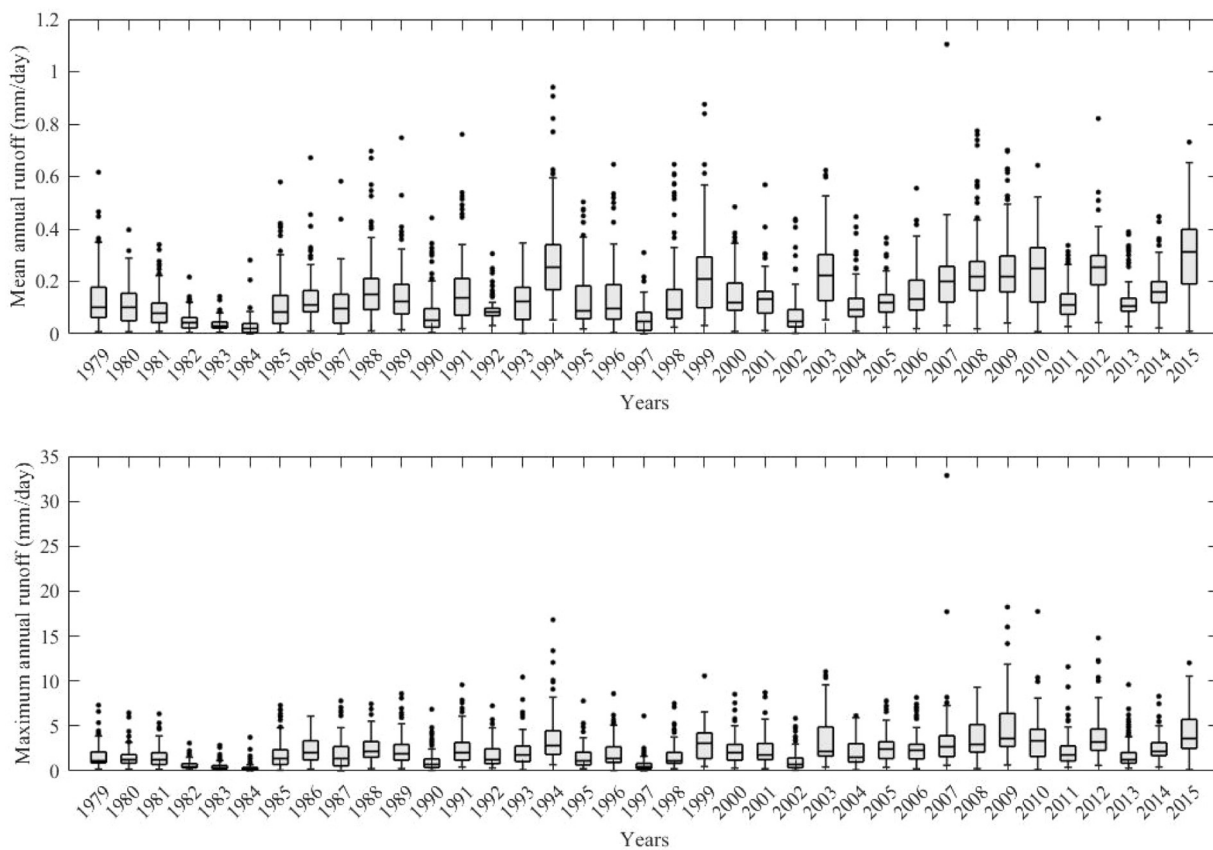


Fig. 2. The boxplots of the mean and maximum annual runoff derived from 116 catchments in the study area.

numbers of clusters, and seven different clustering techniques, the proposed framework leads to 294 various classification results to be compared. The final comparison has been done using the well-known homogeneity measure of Hosking and Wallis (1993, 2005) to respectively determine the best set of explanatory variables for classification, the best clustering approach, and finally the best classification results as the final classified catchments of the study area.

2.2.1. Sets of explanatory variables

Six different sets of explanatory variables are considered in the inter-comparison framework to assess their influences on the classification results and allow us to select the most robust set for catchment classification.

The three sets of variables, based on the collected catchments' physiographic attributes (Table 1), are (1) topographic (TP), (2) land-cover (LU), and (3) the combination of topographic and land-cover (TP+LU) attributes. Also, a set of explanatory variables are generated using the five streamflow signatures proposed by Sawicz et al. (2011), including Runoff Ratio (RR), Baseflow index (BF), Rising Limb Density (RLD), Streamflow Elasticity (SE), and Slope of Flow Duration Curve (SFDC). These metrics have been previously used for the purpose of catchment classification (e.g. Boscarello et al., 2016; Razavi and Coulibaly, 2013; Ssegane et al., 2012).

Runoff ratio, as the ratio of the long-term average streamflow to precipitation, shows the long-term differences between the releases as streamflow and evapotranspiration. The baseflow index is the ratio of the long-term average of annual baseflow to the long-term

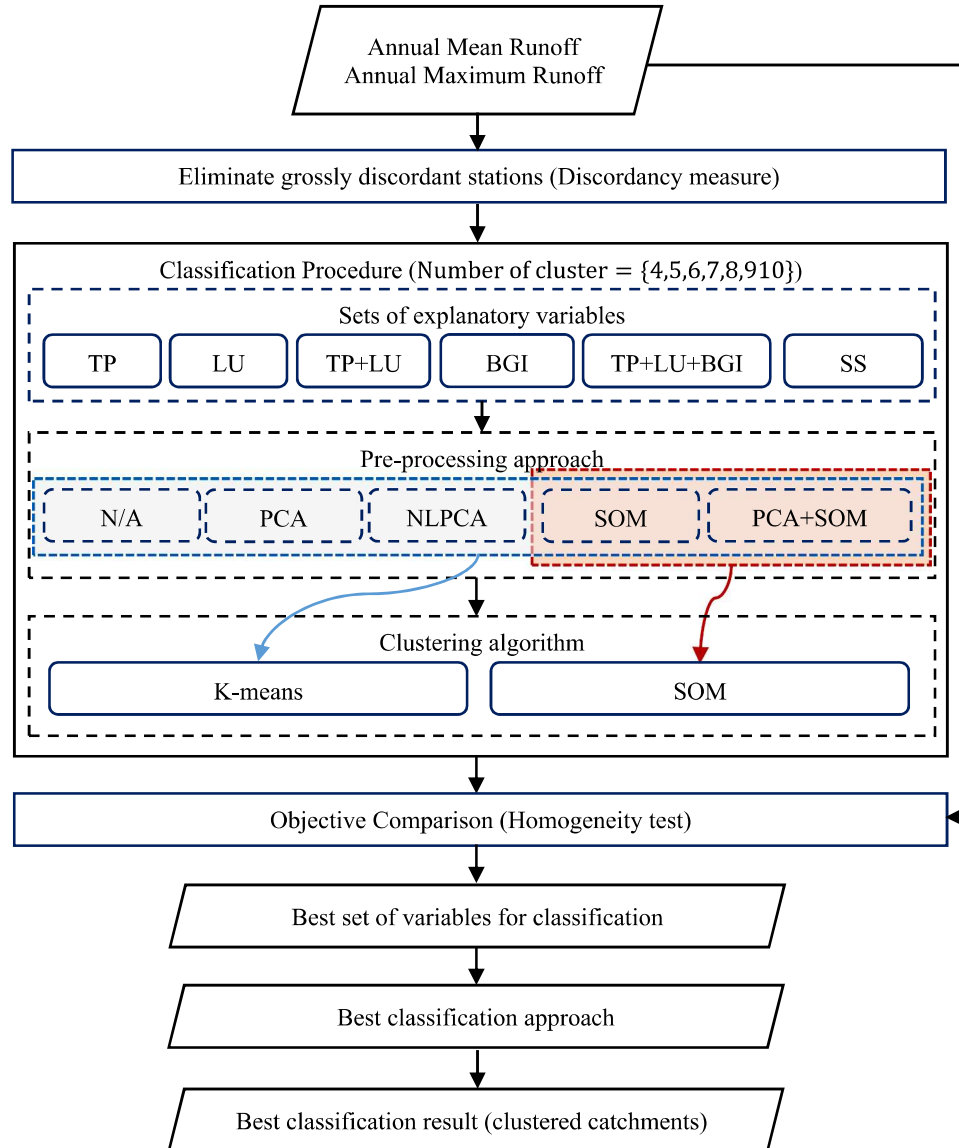


Fig. 3. Flow chart of the applied inter-comparison procedure for catchment classification. Set of explanatory variables are: TP: Topographic attributes, LU: Land-use attributes, BGI: Monthly Bagnouls-Gaussien index, SS: Streamflow signatures. PCA, NLPCA, and SOM are respectively stands for Principal Component Analysis, Non-linear Principal Component Analysis, and Self-Organizing Maps approaches.

average of streamflow and its larger values indicate more contribution of low flow events. Rising limb density represents the flashiness of the hydrograph by comparing the number of actual rising limbs and the total number of times the hydrograph shows increasing flows. Streamflow elasticity shows how sensitive is the catchment streamflow response as a function of changes in input precipitation, and the slope of the flow duration curve, based on daily data, indicates the variability in streamflow events. The equation of all aforementioned streamflow indices and their corresponding ranges in the study area are presented in Table 2. For detailed explanations of the metrics, the readers are referred to the cited references.

Furthermore, the next set of explanatory variables used in this study is based on the Bagnouls-Gausson index (BGI, Bagnouls and Gausson, 1957; Gausson and Bagnouls, 1952) which can be considered as a climate characteristic of the catchments. The monthly BGI is a function of the average monthly temperature (T_i) and precipitation (P_i) being formulated as follows:

$$BGI_i = 2 \times T_i - P_i \quad (1)$$

The positive monthly BGI (i.e. the two times the average monthly temperature is larger than the average monthly precipitation) can be used as an indicator of a dry month (Neira Mendez, 2006). The box plots of the monthly BGI in the study area are presented in Fig. 4. The annual aggregation of the monthly BGI is the well-known aridity index for climate classification. However, to the best of our knowledge, monthly BGI has not been previously implemented for generating hydrologically homogenous regions through catchment classification. Finally, the last set of explanatory variables is the combination of topographic, land-cover, and monthly BGI (TP+LU+BGI) attributes.

2.2.2. Classification techniques

The seven different classification techniques compared in this study are (1) K-means, (2) PCA+K-means, (3) NLPCA+K-means, (4) SOM+K-means, (5) PCA+SOM+K-means, (6) SOM+SOM, and (7) PCA+SOM+SOM, which are developed as a combination of various pre-processing approaches and clustering methods. Principal component analysis (PCA) and non-linear principal component analysis (NLPCA) are the two statistical pre-processing methods that are considered for reducing the dimensionality of the explanatory variables. The K-means algorithm and Self-Organizing Map (SOM) are the two clustering techniques being considered in this study. Besides clustering, SOM is also used as a pre-processing tool for reducing the dimensionality of the data. Implementing SOM as a pre-processing approach before clustering has been previously used in various catchment classification studies (e.g. Boscarello et al., 2016; Farsadnia et al., 2014; Ley et al., 2011; Sharghi et al., 2018). A brief explanation of each method is provided in the following sections.

2.2.2.1. Principal component analysis. The principal component analysis (PCA) is one of the most popular statistical approaches which reduces the dimensionality of the data by linearly transforming a large number of correlated variables into a smaller number of uncorrelated variables, called latent variables or principal components (PCs) (Di Prinzio et al., 2011; Razavi and Coulibaly, 2013). The PCs are derived from an eigenvector decomposition of the covariance matrix of the variables (Aguado et al., 2008) and have the same dimensions as the original data. So, using the matrix of all PCs (Y), the following equation can be used to generate the original sets of variables (X):

$$X = Y \times P^T \quad (2)$$

where P is the loadings (i.e. the eigenvalues) and $(.)^T$ is a transpose function.

By sorting the orthogonal PCs based on their eigenvalues, the first principal component explains the most variability of the original data while the last one accounts for the least variability, so, selecting the first few PCs may be enough to explain the main pattern in the original data (Jackson, 2005). Following Di Prinzio et al. (2011), the threshold of 80 % of the total variance of the original data is used in this study for reducing the dimensionality and selecting the proper number of PCs as the new set of variables for classification.

Table 2
The streamflow signatures and their formulations.

Signatures	Description	Equation	Range
SFDC	Slope of the Flow Duration Curve	$SFDC = \frac{(\ln(Q_{33}) - \ln(Q_{66}))}{0.33}$	3.6 – 24.7
BI	Baseflow Index	$BFI = \frac{Q_B}{Q}$	0.0 – 0.1
RLD	Rising limb density	$RLD = \frac{N_{RL}}{T_R}$	0.2 – 0.5
RR	Runoff Ratio	$RR = \frac{Q}{P}$	0.04 – 2.5
SE	Streamflow Elasticity	$SE = median\left(\frac{dQ\bar{P}}{dP\bar{Q}}\right)$	–1.1 – 4.5

Q_{33} and Q_{66} respectively are the 33rd and 66th percentile of the flow duration curve developed based on daily streamflow data. Q and P are the long-term averages of the daily streamflow and precipitation, respectively, and Q_B is the long-term average of the daily baseflow. $\bar{Q}$ and $\bar{P}$ respectively are the mean annual streamflow and precipitation, and dQ and dP are the inter-annual differences between annual streamflow and annual precipitation respectively. T_R is the number of times the daily hydrograph is rising and N_{RL} is the total number of rising limbs.

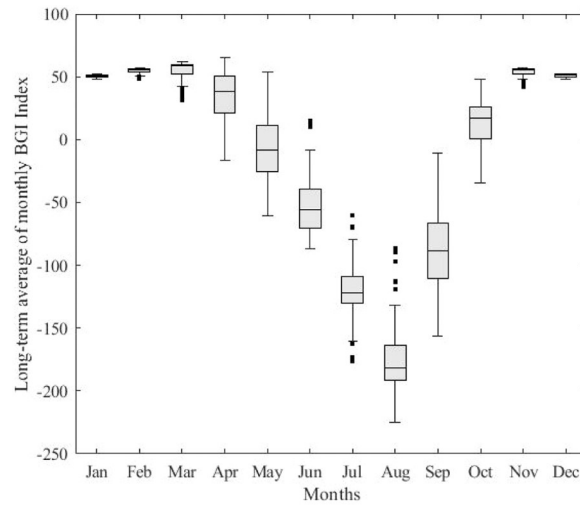


Fig. 4. Box plots of the long-term average of monthly Bagnouls-Gaussien index (BGI) based on various catchments, being used as a set of explanatory variables for catchment classification.

2.2.2.2. Non-linear principal component analysis. Unlike PCA, where the linear correlation between variables is considered, the Non-Linear Principal Component Analysis (NLPCA), proposed by [Kramer \(1991\)](#), is a neural network-based method where non-linear vector functions are used to transform data into a new dataset. In NLPCA, the original data set with m variables (X) is mapped into a feature space (Y) using a vector of n non-linear mapping functions (G):

$$Y = G(X) \quad (3)$$

Then, Y transformed back to the original space (X') using another vector of m non-linear de-mapping functions (H):

$$X' = H(Y) \quad (4)$$

By determining the mapping and de-mapping functions through minimizing the loss of information (e.g. $L = X - X'$) to an acceptable range, Y can be considered as an appropriate representation of the original data set. Therefore, NLPCA is an auto-associative feedforward neural network with three hidden layers of mapping, bottleneck, and de-mapping ([Fig. 5](#)) where the outputs of the bottleneck layer are considered non-linear principal components ([Razavi and Coulibaly, 2013](#)). The mapping and de-mapping layers include nodes with sigmoidal functions which makes the procedure non-linear and selecting the lower number of neurons in the bottleneck layer, compared to mapping and de-mapping layers (i.e. $m < n$), leads to dimension reduction.

In this study, following [Lu and Pandolfo \(2011\)](#) and [Razavi and Coulibaly \(2013\)](#), the two-neuron bottleneck layer gives two non-linear principal components, and the numbers of neurons in mapping and de-mapping layers are set for an optimal fit of the output to the input.

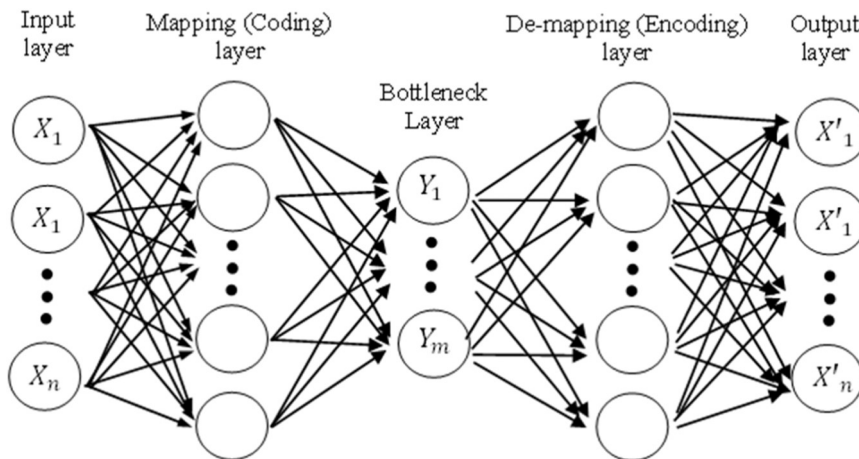


Fig. 5. Schematic diagram of the three-layer neural network used for non-linear principal component analysis.

2.2.2.3. K-means algorithm. K-means algorithm, proposed by MacQueen (1967), is an unsupervised non-hierarchical clustering technique that is commonly used in catchment classification studies (e.g. Farsadnia et al., 2014; Razavi and Coulibaly, 2013; Sharghi et al., 2018). K-means algorithm tries to classify the multivariate data points into k groups by minimizing the intra-cluster distance (i.e. the distance between the data points within each cluster to the centroid of the corresponding cluster). K-means is an iterative algorithm, starting by placing k initial centroids into the data point space and assigning the data points to a group with the closest centroid. Then, the new centroids are recalculated, and the memberships of clusters are changed accordingly. These two steps continue until there is no change in the clusters' centroids (Chang et al., 2008). The Euclidean distance is the most common metric for measuring the distance in this algorithm and it is used in this study. It is of note that the number of clusters, k , should be pre-specified before the analysis. To reduce the effect of initialization and trapping in local minima, as two common concerns of the K-means algorithm (Sharghi et al., 2018), the algorithm runs 50 times, and the best classification result is selected by comparing the Davies-Bouldin (DB) index (Davies and Bouldin, 1979) as one the well-known metrics for evaluating the clustering algorithm:

$$DB = \frac{1}{k} \sum_{i=1, j \neq i}^k \max \left(\frac{\bar{d}_i + \bar{d}_j}{d_{ij}} \right) \quad (5)$$

$\bar{d}_i$ is the intra-cluster distance and d_{ij} is the distance between the centroid of cluster i and j . The smallest value of DB shows the smallest intra-cluster distance while the larger distance between different clusters.

2.2.2.4. Self-Organizing Maps. Self-Organizing Maps (SOM) or Kohonen maps, as it was first introduced by Kohonen in 1982 (Kohonen, 1982), is a neural network-based unsupervised learning procedure for analyzing high dimensional databases. The main goal of SOM is to reduce the dimensionality in a non-linear manner by mapping high-dimensional datasets into a lower-dimensional grid, while it can also be used for clustering data (Razavi and Coulibaly, 2013; Sharghi et al., 2018). As illustrated in Fig. 6, a typical SOM consists of two layers of neurons: (1) the input layer where each variable is represented by a neuron, and (2) the output (Kohonen) layer with a pre-specified number of neurons, mostly arranged in a two-dimensional grid, as a low-dimensional representation of the input layer (Kohonen, 2001). Each neuron in the output layer is connected to all neurons in the input layer through weight vectors which are determined in a competitive, cooperative, and topology-preserving learning procedure. In the training iteration, after arbitrary weights' initialization, for each input vector presented to the model, the distance measure (i.e. Euclidean distance) between the output neurons and the input vector is used to determine the closest neuron as the winning one (best matching unit). Then the weights vector of the winning neurons and their neighbors are adjusted to reproduce the pattern of the input (Sharghi et al., 2018). The SOM training procedure preserves the topology of the data sets by considering the connections between neighboring neurons in the output layer (Di Prinzio et al., 2011). After training, each input vector is assigned its best matching unit in the output layer. Determining the number of output neurons in SOM is a subjective choice of the modeler and can affect the performance of the model. By considering the same number of neurons as the desired number of clusters, SOM can be straightly used for catchment classification (e.g. Di Prinzio et al., 2011; Razavi and Coulibaly, 2013; Toth, 2013) while some studies proposed using a more appropriate size for the SOM output layer and applying a clustering algorithm to classify the output neurons and correspondingly, their associated input vectors (e.g. Boscarello et al., 2016; Farsadnia et al., 2014). One of the optimal numbers of SOM output neurons, suggested by Vesanto

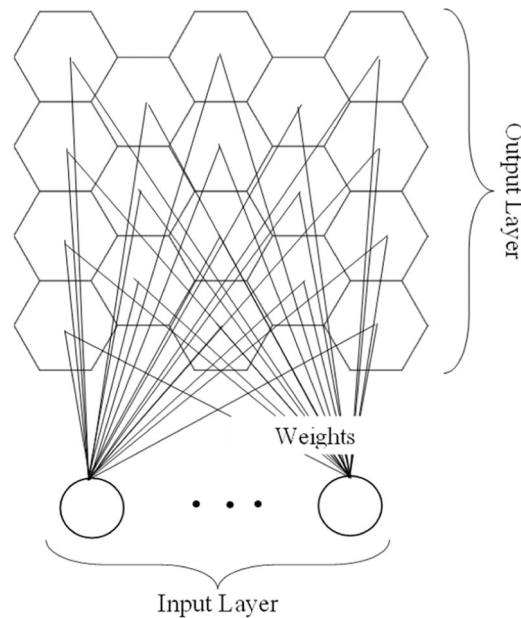


Fig. 6. Schematic diagram of the Self-Organizing Map (SOM).

and Alhoniemi (2000) and commonly used in previous studies, is $5\sqrt{n}$ where n is the number of input vectors.

2.2.3. Discordancy assessment and homogeneity evaluation

The discordancy measure (D), proposed by Hosking and Wallis (1993, 2005), is an L-moment-based metric that can be used to recognize the discordant stations (or catchments) compared to the group. The discordancy measure of catchment i (D_i) is a function of the vector of L-moment ratios, $u_i = [t_2^{(i)}, t_3^{(i)}, t_4^{(i)}]^T$, where $t_2^{(i)}$, $t_3^{(i)}$ and $t_4^{(i)}$ are the sample L-moment ratios (L-Coefficient of Variation (L-CV), L-Skewness, and L-kurtosis, respectively) and formulated as follows (Hosking and Wallis, 2005):

$$D_i = \frac{1}{3} N(u_i - \bar{u})^T A^{-1} (u_i - \bar{u}) \quad (6)$$

$$\bar{u} = \frac{\sum_{i=1}^N u_i}{N} \quad (7)$$

$$A = \sum_{i=1}^N (u_i - \bar{u})(u_i - \bar{u})^T \quad (8)$$

N is the total number of catchments in the group. In general, the catchments with a discordancy measure larger than 3 can be considered discordant (Hosking and Wallis, 1993). The discordancy metric is mainly used to identify and eliminate the grossly discordant catchments, compared to all catchments, as the initial screening procedure (Hosking and Wallis, 1993).

Moreover, Hosking and Wallis (1993, 2005) introduced a homogeneity measure (H_1) for evaluating the homogeneity of the obtained classification result by assuming that except from sampling variability, there is no variation in the sample L-CV of sites in a homogeneous region. The homogeneity metric is determined by comparing the standard deviation of the at-site sample L-CVs with its expected variability, being determined by fitting a four-parameter kappa distribution on the regional average L-CVs of the sites and performing a Monte-Carlo simulation, and can be formulated as follows (Hosking and Wallis, 2005):

$$H_1 = \frac{V_{obs} - \mu_v}{\sigma_v} \quad (9)$$

$$V = \left[\frac{\sum_{i=1}^N n_i (t_2^{(i)} - \bar{t}_2)^2}{\sum_{i=1}^N n_i} \right]^{0.5} \quad (10)$$

V , named V-statistic, is the weighted standard deviation of the L-CVs ($t_2^{(i)}$) and n_i , the weight factor, is the data length in site i . V-statistic can be calculated using the observed data (V_{obs}) and based on the simulated realizations, where their population parameters (i. e. mean and standard deviation), are denoted by μ_v and σ_v , respectively. A larger value of H_1 shows more dispersion in the observed L-moments, which is a sign of heterogeneity in the region. As recommended by (Hosking and Wallis, 1993, 2005), the region is considered homogeneous if $H_1 < 1$, possibly heterogeneous if $1 \leq H_1 < 2$, and definitely heterogeneous if $H_1 \geq 2$. Also, a more relaxed criterion is proposed by Wallis et al., (2007) where the region is considered acceptably homogeneous if $H_1 < 2$ and definitely heterogeneous if $H_1 > 3$. These modified thresholds are proposed to consider the uncertainty derived from measurement errors (Chérif and Bargaoui, 2013). It is of note that the H-statistics can also be calculated using L-Skewness, and L-kurtosis (denoted by H_2 and H_3 , respectively), however, compared to the H_1 statistic, they lack the power to properly discriminate between heterogeneity and homogeneity and H_1 is suggested as the principal metric for evaluating homogeneity (Farsadnia et al., 2014; Hosking and Wallis, 1993, 2005; Sharghi et al., 2018).

For comparing different classification results, four different performance criteria are defined as a function of the calculated H_1 statistics for different clusters:

$$H_{MEAN} = \frac{\sum_{i=1}^k H_{1i} \times m_i}{N_T} \quad (11)$$

$$H_{R1} = \frac{\sum_{i=1}^k I_1(H_{1i}) \times m_i}{N_T} \quad (12)$$

$$H_{R2} = \frac{\sum_{i=1}^k I_2(H_{1i}) \times m_i}{N_T} \quad (13)$$

Where m_i is the size of the cluster i and N_T is the total number of classified catchments and the I function is defined as follows:

$$I_j(x) = \begin{cases} 1 & \text{if } x < j \\ 0 & \text{if } x \geq j \end{cases} \quad (14)$$

For each classification result, H_{MEAN} is the average of the H_1 statistics of the corresponding clusters weighted by the size of the clusters. Similar to H_1 , the lower value of H_{MEAN} illustrates the better classification results in terms of developing homogeneous regions. H_{R1} and H_{R2} are the homogeneity ratios considering the homogeneity thresholds of 1 and 2, respectively. By considering the size of homogeneous and heterogeneous regions, these two metrics are proper metrics for comparing the homogeneity of different classification results based on the suggested thresholds in the literature (Ahani et al., 2020). H_{R1} and H_{R2} vary between 0 and 1 with the best value of 1 indicating that all regions are homogeneous with respect to the considered threshold. Additionally, the median of the H_1 statistics (H_{MED}) is used as a performance metric. This is calculated by assigning the same H_1 value to all members of a cluster and then determining the median across all members. This approach provides an alternative measure of central tendency while accounting for

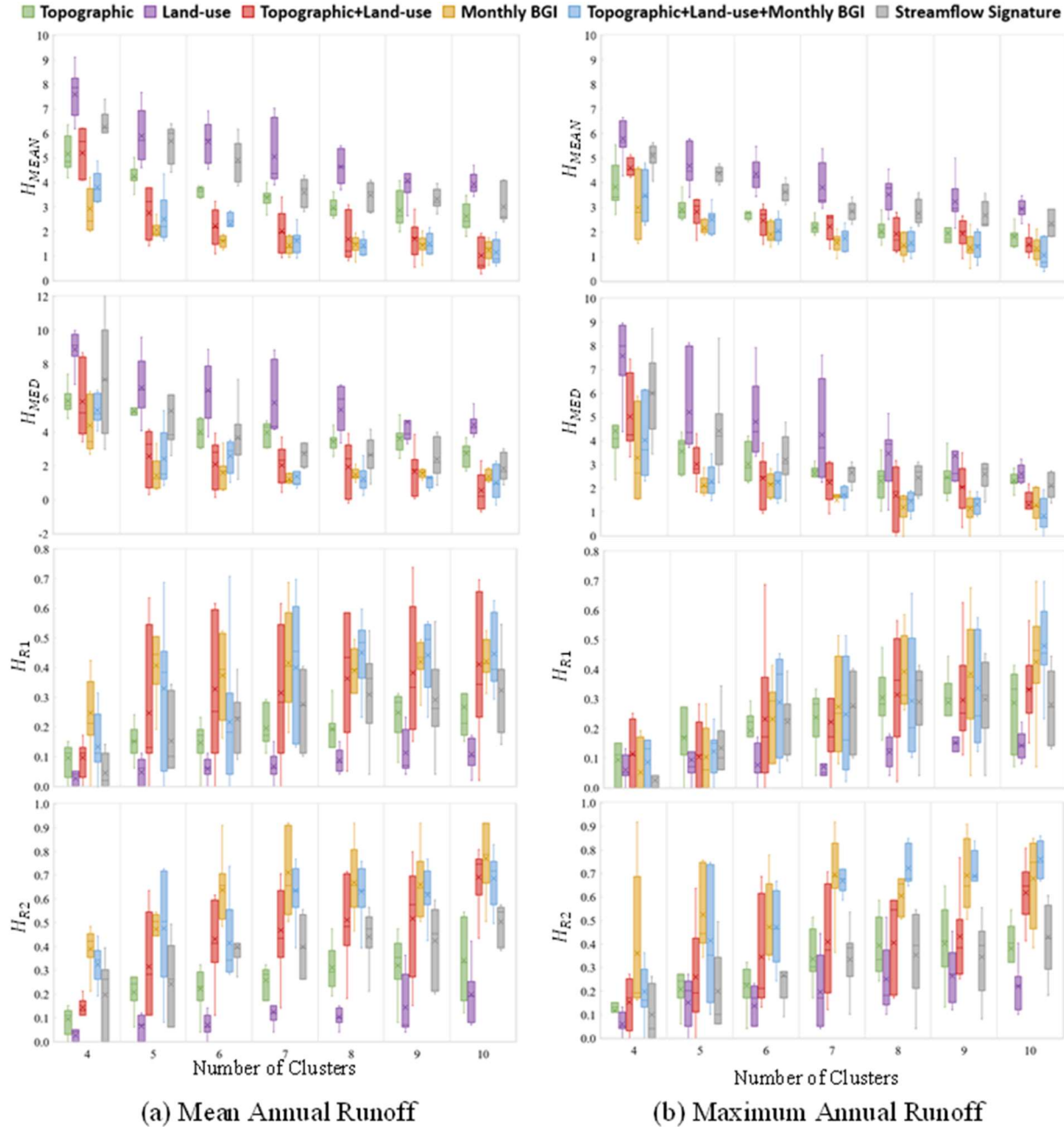


Fig. 7. Box plots of different homogeneity performance criteria of the classified catchments based on (a) mean annual runoff and (b) maximum annual runoff by implementing various explanatory variables, derived from seven different clustering approaches.

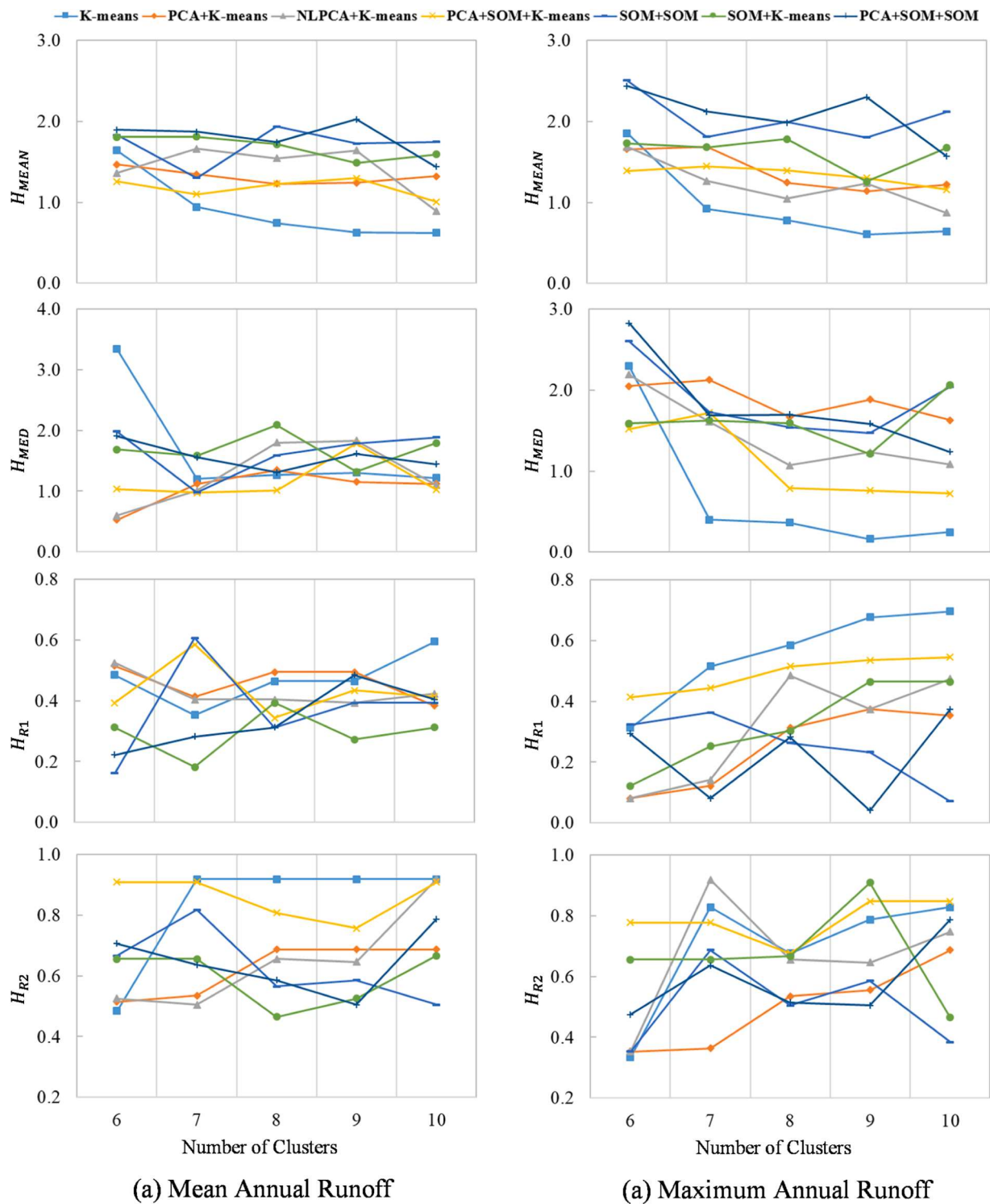


Fig. 8. Comparison of the performance of different clustering approaches with different numbers of clusters, using monthly BGI index as the set of explanatory variables, based on (a) mean and (b) maximum annual runoff.

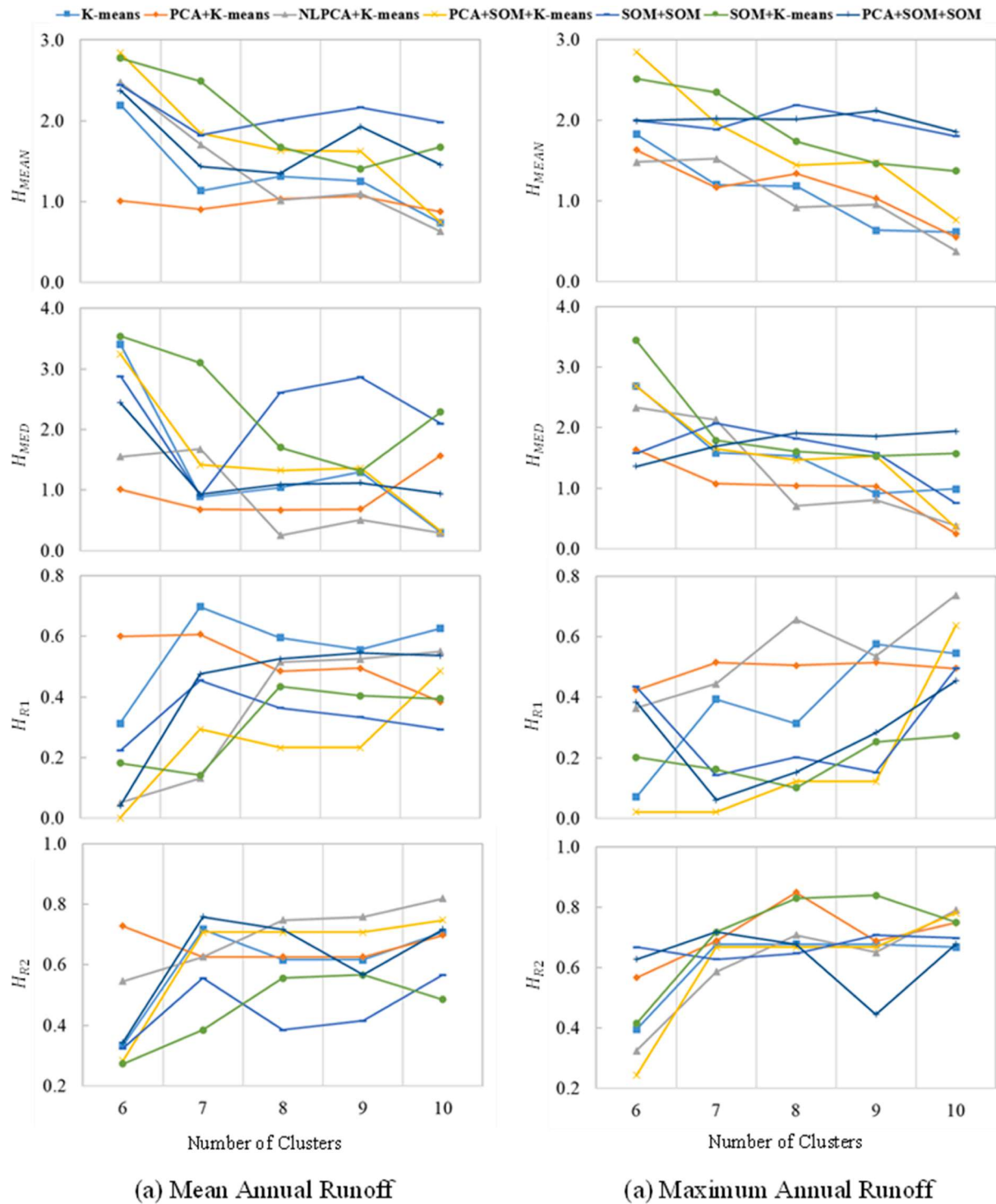


Fig. 9. Comparison of the performance of different clustering approaches with different numbers of clusters, using monthly BGI index in combination with Topographic and land-use attributes as the set of explanatory variables, based on (a) mean and (b) maximum annual runoff.

the size of each cluster.

3. Results and discussions

3.1. Evaluating the effects of explanatory variables on classification results

After eliminating the 17 highly discordant catchments, which are determined by discordancy assessment based on both mean and maximum annual runoff time series as the first step of the classification procedure, 99 non-discordant catchments in the study area remained to be classified using 294 different classification configurations (Fig. 3). As previously mentioned, selecting the proper set of variables for catchment classification is the prime step for possessing reliable and homogeneous classified regions. Therefore, as the first step, the comparison is made between the classification results of various clustering procedures using the six different sets of explanatory variables to assess the effects of variable selection on classification results (Fig. 7). In this figure, the box plots for each set are generated using the results of seven different clustering procedures. The first thing that stands out from the results is that implementing the monthly BGI, as the combination of two main climate characteristics, leads to relatively better classification results in terms of predicting both mean and maximum annual runoff. Either being used alone as the explanatory variables or in conjunction with topographical and land-use attributes, all performance metrics, especially the H_{R2} , show better classification results compared to the other four sets of variables (Fig. 7). This confirms that the climate properties have the dominant role in controlling the hydrological behaviors of the catchments (Sawicz et al., 2011). Moreover, the results show that implementing land-cover attributes, alone, as the explanatory variables for classification do not lead to hydrologically homogenous regions in the case of considering both annual mean and maximum runoff. This is in line with previous studies showing that land-use attributes could not be a sufficient source for defining catchment hydrological behavior (Merz and Blöschl, 2009). On the other hand, however, the comparison of implementing topographic attributes and the combination of the topographic and land-use attributes for catchment classification shows that using land-cover data besides topographical characteristics of the catchments can enhance the results and lead to relatively more homogenous clusters. These enhancements are more pronounced in the case of considering mean annual runoff as the variable of interest, compared to maximum annual runoff. Moreover, interestingly, no advantage can be seen in using the streamflow signatures as the explanatory variables. Although the used streamflow indices are defined in a way to capture most of the catchments' hydrological behavior (Razavi and Coulibaly, 2013; Ssegane et al., 2012), the classification based on them does not outperform the classification based on BGI or the combination of topographic and land-cover attributes.

3.2. Comparison of clustering methods

Considering monthly BGI and the combination of monthly BGI with the topographic and land-use attributes, which are found as the two best sets of explanatory variables for the catchment classification in the study area, the seven clustering algorithms, presented in Section 2.2.2, are compared. Fig. 8 illustrates the comparison between the performances of the classification results derived from different clustering approaches considering various numbers of clusters based on monthly BGI as the explanatory variable. The comparison of H_{MEAN} measurements show the clear advantage of the K-means algorithm in producing homogenous regions in terms of both mean and maximum annual flows, compared to the other more complex approaches. Except for the number of clusters equal to six, K-means lead to the best H_{MEAN} in all other cases with at least 25 % and 15 % improvements compared to the second-best method in the case of considering maximum and mean annual flows, respectively. Also, both homogeneity ratios (H_{R1} and H_{R2}) confirm the reliable performance of the K-means clustering algorithm. Considering the maximum annual runoff as the variable of interest, H_{R1} measurements show that K-means produce the most homogeneous regions and PCA+SOM+K-means is in the second place with a noticeable distance. Moreover, the outperformance of K-means and PCA+SOM+K-means methods for generating hydrological

Table 3

The homogeneity metrics of the best classification results derived from various classification methods and the monthly BGI as the explanatory variables based on mean and maximum annual runoff. Bold values represent the best performance, while underlined values indicate the worst.

Variable of interest	Classification method	Best Number of Clusters	Homogeneity metrics			
			H_{MEAN}	H_{MED}	H_{R1}	H_{R2}
Annual Mean Flow	K-means	10	0.62	1.22	0.60	0.92
	PCA+K-means	9	1.24	1.15	0.49	0.69
	NLPCA+ K-means	10	0.89	1.10	0.42	0.92
	PCA+SOM+K-means	7	1.10	0.97	0.59	0.91
	SOM+SOM	7	1.30	0.98	0.60	0.82
	SOM+ K-means	10	<u>1.59</u>	<u>1.80</u>	<u>0.31</u>	<u>0.67</u>
	PCA+SOM+SOM	10	1.44	1.44	0.40	0.79
Annual Maximum Flow	K-means	10	0.64	0.24	0.70	0.83
	PCA+K-means	10	1.22	1.63	<u>0.35</u>	<u>0.69</u>
	NLPCA+ K-means	10	0.87	1.08	0.47	0.75
	PCA+SOM+ K-means	10	1.16	0.72	0.55	0.85
	SOM+SOM	7	<u>1.81</u>	<u>1.73</u>	0.36	<u>0.69</u>
	SOM+ K-means	9	1.26	1.21	0.46	0.91
	PCA+SOM+SOM	10	1.57	1.24	0.37	0.79

homogeneous regions for predicting mean annual runoff is illustrated using H_{R2} criteria. Also, besides PCA+SOM+K-means and PCA+K-means, the K-means algorithm is among the most reliable models based on the H_{R1} of mean annual runoff and H_{R2} of the maximum annual runoff criteria.

On the other hand, Fig. 9 represents the performances of the classification results derived from different clustering approaches based on using the combination of monthly BGI, topographic, and land-use attributes. The results do not show a clear tendency toward the outperformance of one specific approach. However, in general, based on all four performance criteria, the approaches including K-means clustering method (e.g. K-means, PCA+K-means, and NLPCA+K-means) lead to relatively more homogenous classification compared to SOM-based approaches in terms of both maximum and mean annual runoff. Considering the mean annual runoff as the variable of interest, H_{R1} metric shows a slight superiority of the K-means approach and H_{R2} measurements illustrate the relatively better performance of NLPCA+K-means and PCA+K-means. Moreover, while no clear differences can be found looking at the H_{R2} ratio for maximum annual runoff, the H_{R1} ratio shows the reliable performance of K-means, NLPCA+K-means, and PCA+K-means for generating hydrological homogeneous regions.

To complete the comparison, the best number of clusters is selected separately for each clustering algorithm and the comparison is made between them. We chose the best number of clusters by visually comparing all three criteria (H_{MEAN} , H_{R1} and H_{R2}). Comparing the calculated homogeneity metrics of the final selected classification results based on the monthly BGI explanatory variable (Table 3) confirms the relative superiority of the K-means method in dividing catchments into homogenous regions, compared to the other classification methods. Considering the mean annual runoff, the K-means algorithm leads to the best H_{MEAN} metric and is among the best methods comparing both homogeneity ratios (i.e. H_{R1} and H_{R2}). The measurements based on the annual maximum runoff also show the superior performance of the K-means algorithm. Except for H_{R2} , which shows the competitive performance of K-means, SOM+K-means, and PCA+SOM+K-means. The relative superiority of the K-means algorithm, compared to the second-best approach, in terms of H_{MEAN} and H_{R1} is more than 25 % (K-means achieved H_{MEAN} of 0.64 and H_{R1} of 0.70 compared to the second-best H_{MEAN} value of 0.87 (from NLPCA+K-means) and H_{R1} value of 0.55 (from PCA+SOM+K-means), respectively). Furthermore, contrary to previous findings suggesting the advantages of employing a pre-processing approach before catchment classification (e.g. Di Prinzio et al., 2011; Farsadnia et al., 2014; Ssegane et al., 2012), the comparison of K-means with NLPCA+K-means, PCA+K-means, and SOM+K-means indicates that implementing a pre-processing approach to reduce the dimensionality of monthly BGI does not enhance the results. This lack of improvement may be attributed to the loss of essential information for clustering caused by dimensionality reduction. Therefore, BGI values could be simply used with K-means which simplifies the classification process without loss of performance.

Moreover, the performances of the final selected classification results for each clustering algorithm considering the combination of monthly BGI, topographic, and land-use attributes are presented in Table 4. In line with the previous conclusions, the comparison shows the higher capability of the K-means-based clustering algorithms in producing hydrologically homogenous regions. Although the K-means algorithm leads to relatively good results, implementing the NLPCA or PCA preprocessing approach before the K-means clustering analysis shows better performance. Focusing on mean annual runoff, except for H_{R1} metric which shows the advantage of the sole implementation of K-means approach, the other metrics illustrate the superior performance of NLPCA+K-means. Considering annual maximum runoff as a variable of interest also leads to the same conclusion where NLPCA+K-means outperforms the other approaches. The improvement due to implementing NLPCA preprocessing approach is in line with previous studies which recommend reducing the dimensionality when there is a high number of variables used for classification.

Overall, comparing the implementation of K-means along with the monthly BGI versus the NLPCA+K-means in conjunction with the combination of the monthly BGI, topographic, and land-use attributes do not show significant differences. However, K-means with the monthly BGI shows slightly better performance compared to the NLPCA+K-means with all spatial variables and the monthly BGI, and this can be justifiable by the presence of correlation between most of the topographic and land-use attributes with the monthly BGI

Table 4

The homogeneity metrics of the best classification results derived from various classification methods and the combination of topographic, land-use, and BGI attributes as the explanatory variables based on mean and maximum annual runoff. Bold values represent the best performance, while underlined values indicate the worst.

Variable of interest	Classification method	Best Number of Clusters	Homogeneity metrics			
			H_{MEAN}	H_{MED}	H_{R1}	H_{R2}
Annual Mean Flow	K-means	10	0.73	0.32	0.63	0.69
	PCA+K-means	7	0.90	0.68	0.61	0.61
	NLPCA+ K-means	10	0.63	0.29	0.55	0.80
	PCA+SOM+K-means	10	0.74	0.32	0.49	0.73
	SOM+SOM	7	<u>1.82</u>	0.90	0.46	<u>0.54</u>
	SOM+ K-means	9	1.40	<u>1.31</u>	<u>0.40</u>	0.55
	PCA+SOM+SOM	7	1.43	0.93	0.48	0.74
Annual Maximum Flow	K-means	9	0.64	0.91	0.58	0.70
	PCA+K-means	10	0.55	0.25	0.50	0.75
	NLPCA+ K-means	10	0.38	0.38	0.65	0.79
	PCA+SOM+ K-means	10	0.76	0.35	0.64	0.78
	SOM+SOM	10	1.80	0.75	0.50	0.70
	SOM+ K-means	10	1.37	1.57	<u>0.27</u>	0.75
	PCA+SOM+SOM	10	<u>1.86</u>	<u>1.94</u>	0.45	<u>0.68</u>

index.

3.3. Final classified catchments of the region

The final catchment classification results based on using monthly BGI as the most consistent explanatory variable, in conjunction with the K-means algorithm, which is recognized as the most robust approach for clustering catchments in the selected Sahelian watersheds, are shown in Fig. 10. Also, Fig. 11 represents the comparison of the average monthly BGI, precipitation, and temperature for each cluster. The region from north to south is divided into ten groups consistent with the regional precipitation increasing regime change (Fig. 1) which is also confirmed by comparing the average of the monthly precipitation for each cluster (Fig. 11). Cluster 1 located in the northern area of the region covers the relatively driest and warmest area while the highest precipitation occurs in clusters 9 and 10 which are in the southern part of the study area.

Moreover, for a better comparison of the clusters, the normalized ranges of the geospatial attributes of the catchments (both topographic and land-use) in each cluster are illustrated in Fig. 12. It can be recognized that the clusters located in the northern half of the region with higher latitude and longitude (i.e. Clusters 1, 2, 4, and 6) consist of catchments with relatively lower elevation and slope while Clusters 7, 8, 9, and 10 covers the elevated and highly sloped catchments in the wet southern part of the region. In terms of land cover, as presented in Table 1, Cluster 1 includes catchments in the northern area characterized by the highest portion of bare land and grassland and relatively lower agricultural usage. The tree land cover is increased from north to south leading to the clusters of catchments with more tree coverage and deeper root zone depths in the western heights (i.e. clusters 9 and 10). On the other hand, Clusters 3 and 4 have the relatively highest agricultural usage with the lower root depths (Sawicz et al., 2011). It is worthy of note that the variation of water, wetland, and urban land covers of the catchments are small and can be ignored in the comparison process.

Furthermore, the scatter plots of the first and second l-moments (L-coefficient of variation and L-skewness) for both mean annual runoff and maximum annual runoff (Fig. 13) provide insights into the hydrologic behavior of catchments within each cluster. Cluster 1 shows moderate variability and skewness for mean annual runoff, indicating relatively stable hydrologic behavior, while demonstrates low variability in maximum annual runoff, suggesting it experiences less extreme flow events. Cluster 6 features catchments with relatively higher L-skewness and L-coefficient of variation indicating the significant impact of extreme events on its annual runoff. In contrast, clusters 8 and 9, including catchments in the southern parts of the region, demonstrate stable hydrologic behavior with low variability and skewness and are less prone to hydrologic extremes. Similarly, cluster 3 has low variability and skewness considering

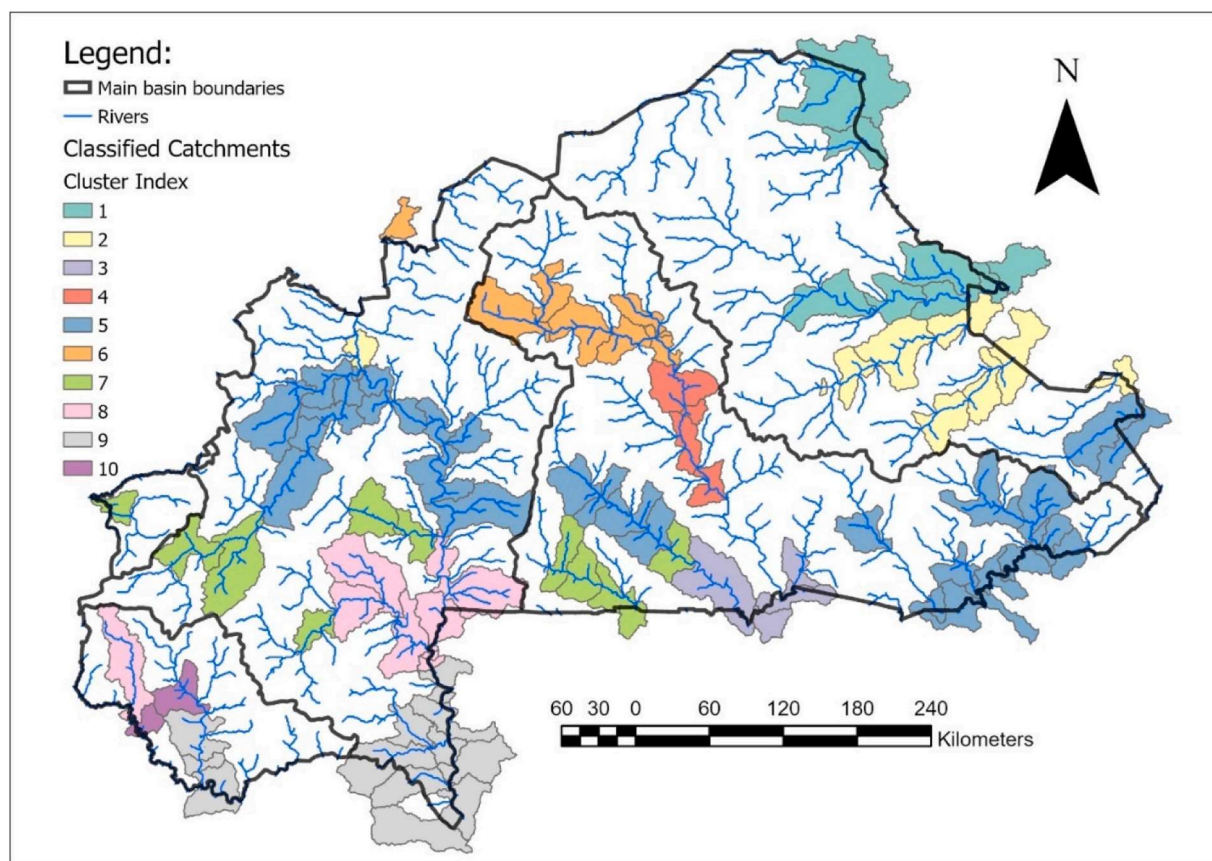


Fig. 10. Classified catchments being identified using the monthly BGI index and K-means algorithm.

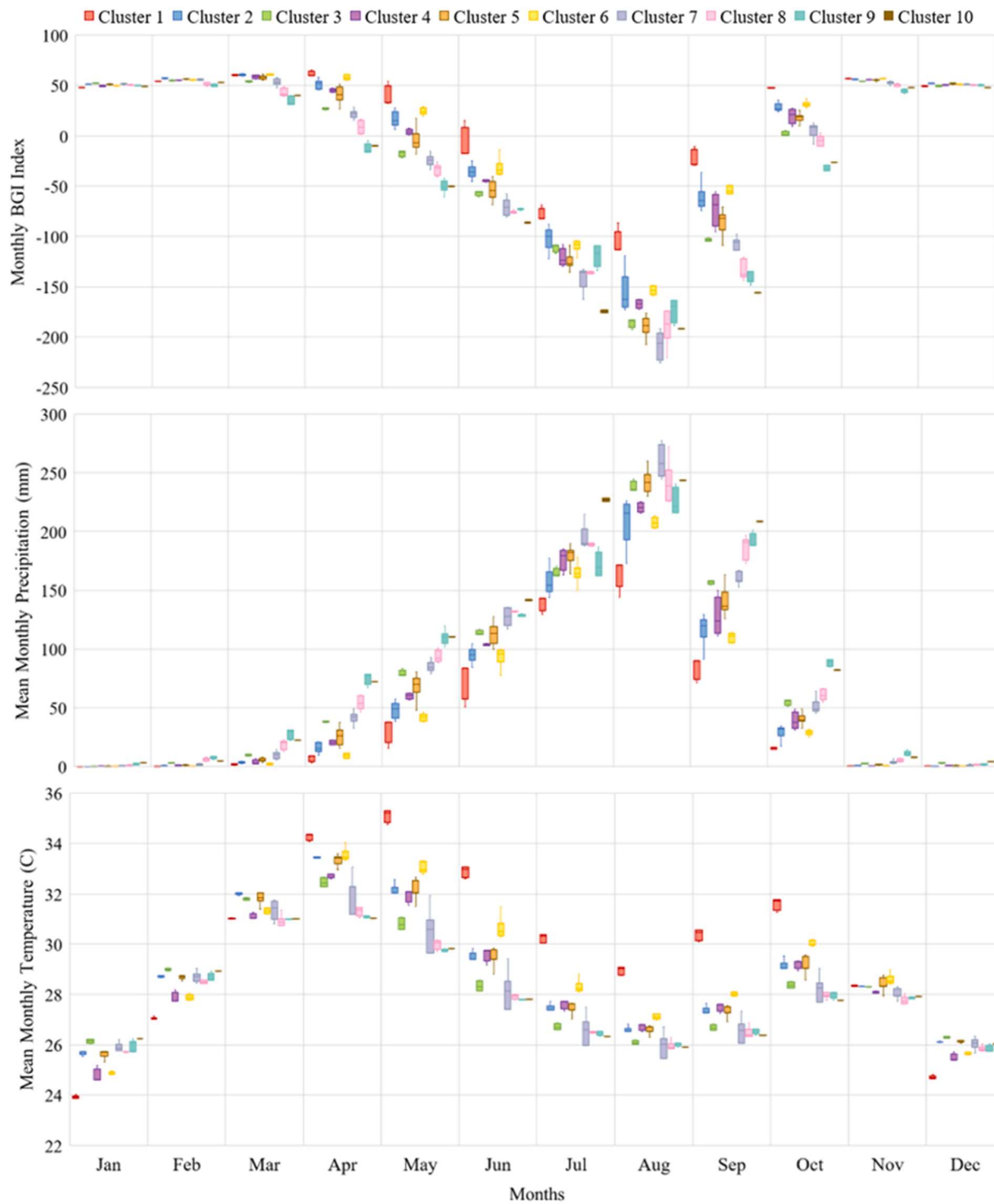


Fig. 11. The box plots of the monthly BGI, average monthly temperature, and average monthly precipitation for each cluster (from 1 to 10) of the final classification results.

maximum annual runoff while exhibiting moderate variability in terms of mean annual runoff. On the other hand, Cluster 7, shows relatively low variability and skewness considering average runoff, while in terms of peak values, it has relatively high variation with moderate skewness. Furthermore, the removal of discordant sites is justified by their distinctly different hydrologic behavior compared to other catchments as shown in Fig. 13. These discordant sites exhibit distinct variation and skewness in terms of either mean or

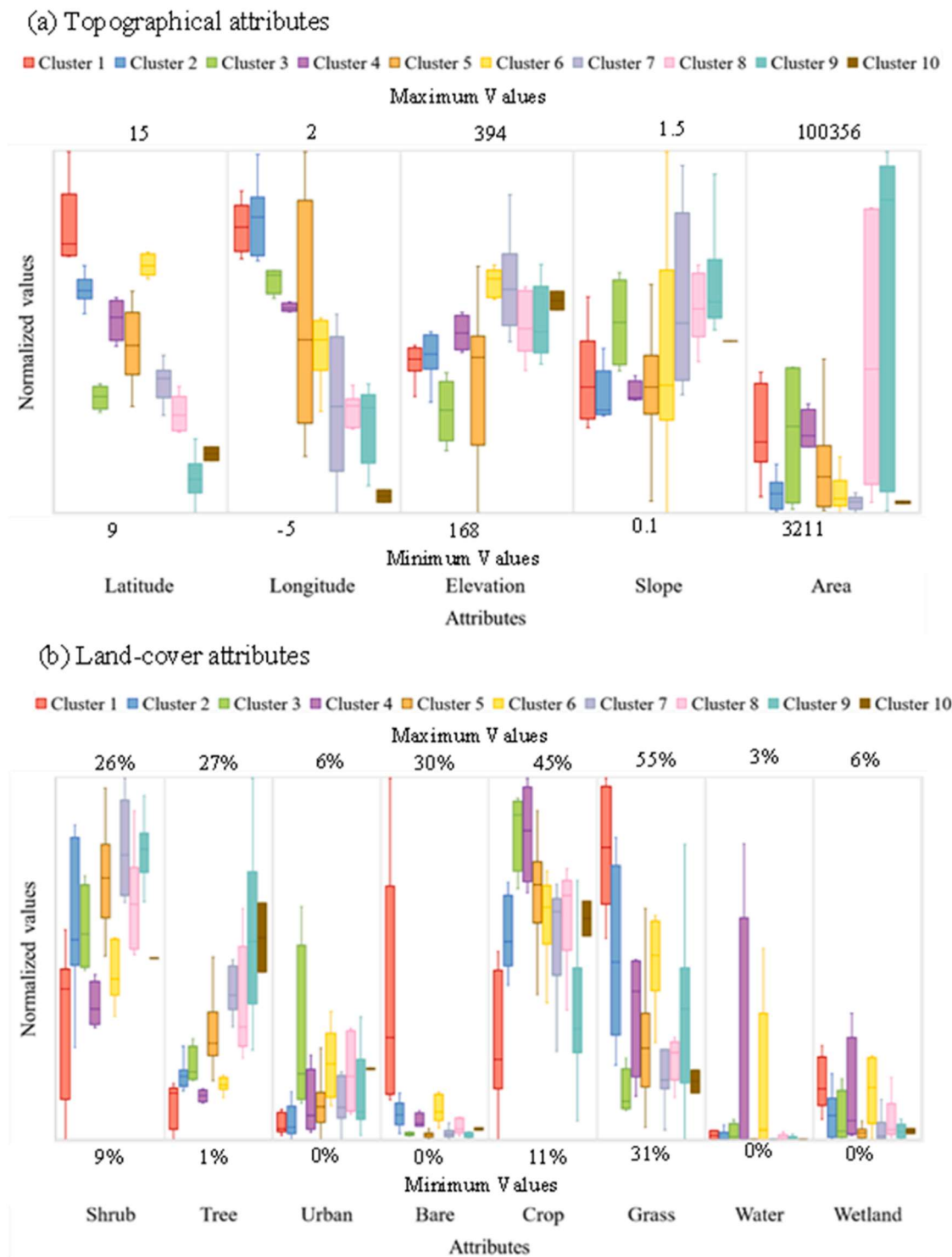


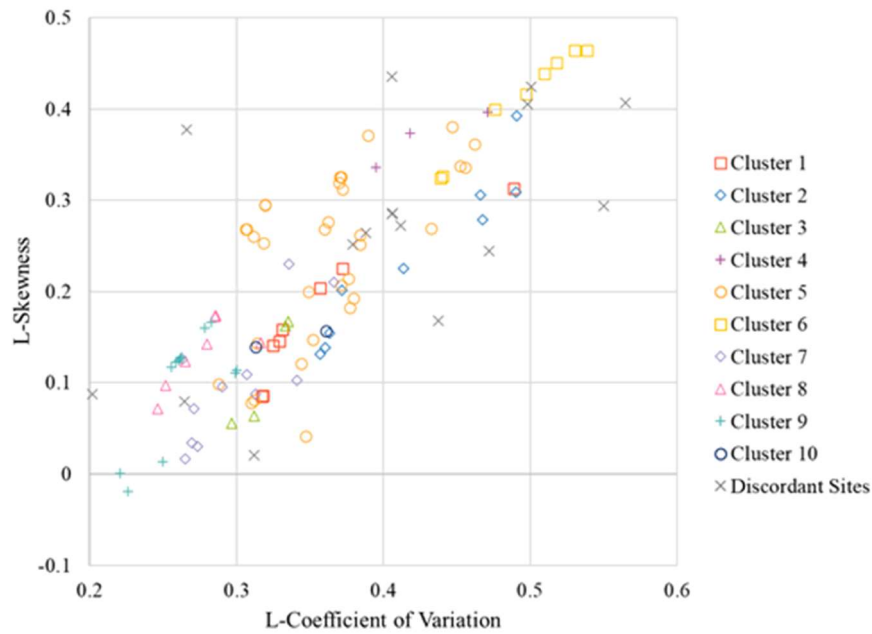
Fig. 12. The box plots of the normalized topographic and land-use attributes of individual clusters.

maximum annual runoff and therefore considering them can negatively affect the comparison results.

4. Summary and conclusions

This study proposed an inter-comparison framework for evaluating the ability of various classification techniques in combination

(a) Mean Annual Runoff



(b) Maximum Annual Runoff

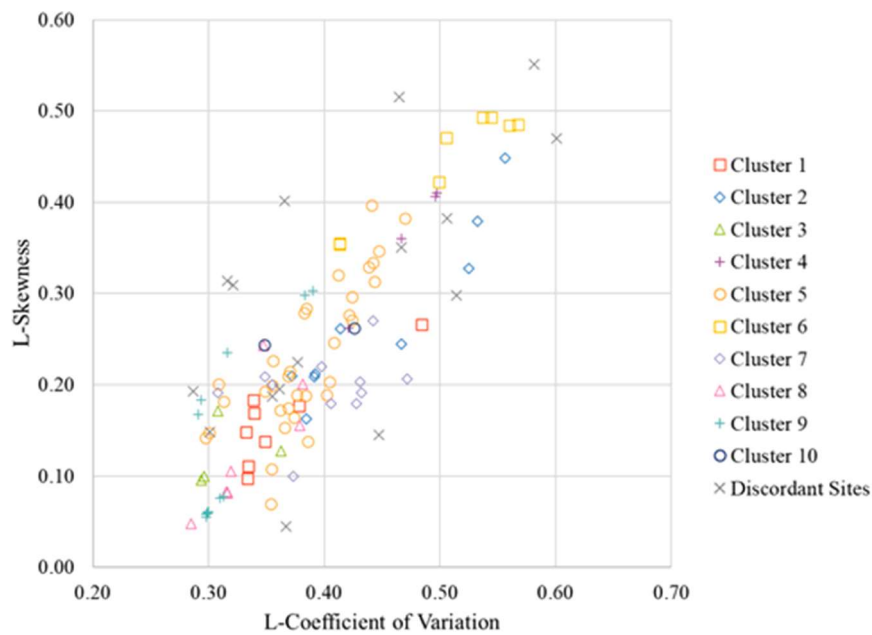


Fig. 13. First (L-Coefficient of Variation) and Second (L-Skewness) l-moment ratios of catchments in each clusters based on (a) mean and (b) maximum annual runoff.

with different sets of explanatory variables for catchment classification purposes. Using the proposed framework, the combination of six different sets of explanatory variables (i.e. topographic attributes, land-use attributes, combined topographic and land-use attributes, monthly BGI index, the combination of monthly BGI, topographic, and land-use attributes, and streamflow signatures) and seven different classification approaches, being developed using various pre-processing methods (PCA, NLPCA, and SOM) and clustering algorithms (K-means and SOM), are compared to identify the homogeneous regions in Sahelian watersheds located in Burkina Faso, West Africa. The comparison is made by considering different numbers of clusters to make the results more robust. Four different

performance criteria developed based on the homogeneity test of Hosking and Wallis (2005) were used to comprehensively compare the classification results using annual mean and annual maximum runoff as two variables of interest.

The summary of the results from applying the proposed framework in the region of interest is as follows:

- Comparing different sets of explanatory variables, in general, shows that using the monthly BGI, either alone or in combination with topographic and land-use attributes, leads to the best classification results in terms of homogeneity. The BGI is a combination of temperature and precipitation as the two main climatology variables. This finding aligns with earlier studies indicating that the dominant characteristic that controls the catchments' hydrological behaviors is climate properties.
- Although in agreement with previous findings, the results indicate that the sole use of land-cover data as the explanatory variables for identifying hydrologically homogeneous regions is not recommended, but its combination with topographic attributes can enhance the quality of the catchment classification results.
- Comparing different classification techniques using monthly BGI as the explanatory variable surprisingly indicates that despite its less advanced structure, the K-means clustering algorithm is consistently superior to the other more advanced methods in terms of identifying hydrologically homogeneous regions. However, PCA+SOM+K-means provides competitive performance.
- The comparison of different classification techniques using the combination of monthly BGI, topographic, and land-use attributes as the explanatory variables shows that methods including K-means clustering perform better and implementing the NLPKA preprocessing approach for reducing dimensionality enhance the results.
- The most homogeneous clustering of the selected catchments was formed with ten groups by using the K-means clustering algorithm based on the monthly BGI.

Overall, the superior performance of considering the monthly BGI index indicates its potential to be used as the explanatory variable for the classification of ungauged watersheds, which will be particularly useful for the poorly gauged Sahelian watersheds. The BGI index is only based on temperature and precipitation and can be comfortably estimated nowadays using various sources of global climate or reanalysis data which makes it a good candidate for data-scarce regions. However, it is important to acknowledge that these meteorological data products often have limitations, such as requiring bias correction, and cannot always be utilized without addressing these issues. Moreover, another interesting implication of the results is the superiority of the K-means algorithm in providing hydrologically homogeneous catchments based on the monthly BGI index. K-means is a well-known and easily accessible approach that can be effortlessly used not only by scientists but also by hydrologists and water resources engineers worldwide.

Using the combination of the K-means algorithm and the monthly BGI can potentially be considered a reliable tool for the purpose of regionalization in ungauged or poorly gauged basins. However, for more comprehensive conclusions, further application of the proposed inter-comparison framework in study regions with different climatology and geography is recommended. Although this study proposed an adaptable comparison framework that can be applied in other regions, the search for a universally robust classification system continues and it is uncertain whether the combination of BGI and K-means could work robustly across the globe, especially for regions with less local-scale variability in climatic indices.

Furthermore, the study did not address the inherent uncertainties in the explanatory variables and the variables of interest, and future studies should consider incorporating uncertainty analysis into the comparison framework to provide more comprehensive results. Moreover, there are other clustering techniques and catchment attributes that are worth including in the comparison procedure in future studies. In addition to the selection of explanatory variables and clustering algorithms, the evaluation of the robustness of the classification techniques by assessing the degree of consistency of different classes (Wolfe et al., 2019) is worth to be added in the future modification of the intercomparison framework. This is an important consideration that could help explore the degree to which the classification is robust.

CRediT authorship contribution statement

Pedram Darbandsari: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization, Writing – review & editing. **Paulin Coulibaly:** Conceptualization, Methodology, Validation, Resources, Writing – review & editing, Visualization, Supervision, Funding acquisition. **Jafet C.M. Andersson:** Validation, Resources, Writing – review & editing, Visualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by the Natural Science and Engineering Research Council (NSERC) of Canada, grant NSERC Canadian FloodNet [NETGP-451456]. The authors are grateful to the Swedish Meteorological and Hydrological Institute (SMHI) for providing the HYPE model data (<http://hypeweb.smhi.se/>), and to the “Direction Générale des Ressources en Eau (DGRE)” and the “Agence Nationale de la Météorologie (ANAM)” of Burkina Faso for providing the hydrological and meteorological data. The authors acknowledge the contribution of Dr. F. Awol and Dr. J. Leach in the pre-processing of the HYPE model outputs and in generating

monthly BGI values for the study area.

Data availability

Data will be made available on request.

References

- Abdolhay, A., Saghaian, B., Soom, M.A.M., Ghazali, A.H.B., 2012. Identification of homogenous regions in Gorganrood basin (Iran) for the purpose of regionalization. *Nat. Hazards* 61 (3), 1427–1442. <https://doi.org/10.1007/s11069-011-0076-2>.
- Aguado, D., Montoya, T., Borrás, L., Seco, A., Ferrer, J., 2008. Using SOM and PCA for analysing and interpreting data from a P-removal SBR. *Eng. Appl. Artif. Intell.* 21 (6), 919–930. <https://doi.org/10.1016/j.engappai.2007.08.001>.
- Ahani, A., Mousavi Nadoushani, S.S., Moridi, A., 2020. Regionalization of watersheds by finite mixture models. *J. Hydrol.* 583, 124620. <https://doi.org/10.1016/j.jhydrol.2020.124620>.
- American Society of Civil Engineers. (1996). *Hydrology Handbook*. American Society of Civil Engineers. <https://doi.org/10.1061/9780784401385>.
- Arheimer, B., Pimentel, R., Isberg, K., Crochemore, L., Andersson, J.C.M., Hasan, A., Pineda, L., 2020. Global catchment modelling using World-Wide HYPE (WWH), open data and stepwise parameter estimation. *Hydrol. Earth Syst. Sci.* 24, 535–559. <https://doi.org/10.5194/hess-24-535-2020>.
- Bagnouls, F., Gaussen, H., 1957. Les climats biologiques et leur classification. *Ann. De. G. éographie* 66 (355), 193–220. <https://doi.org/10.3406/geo.1957.18273>.
- Berg, P., Almén, F., Bozhinova, D., 2021. HydroGFD3.0: a 25 km global near real-time updated precipitation and temperature data set. *Earth Syst. Sci. Data* 13, 1531–1545. <https://doi.org/10.5194/essd-13-1531-2021>.
- Boscarello, L., Ravazzani, G., Cislighi, A., Mancini, M., 2016. Regionalization of Flow-Duration Curves through Catchment Classification with Streamflow Signatures and Physiographic–Climate Indices. *J. Hydrol. Eng.* 21 (3), 05015027. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001307](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001307).
- Burn, D.H., Goel, N.K., 2000. The formation of groups for regional flood frequency analysis. *Hydrol. Sci. J.* 45 (1), 97–112. <https://doi.org/10.1080/02626660009492308>.
- Chang, F.-J., Tsai, M.-J., Tsai, W.-P., Herricks, E.E., 2008. Assessing the ecological hydrology of natural flow conditions in Taiwan. *J. Hydrol.* 354 (1), 75–89. <https://doi.org/10.1016/j.jhydrol.2008.02.022>.
- Chérif, R., Bargaoui, Z., 2013. Regionalisation of Maximum Annual Runoff Using Hierarchical and Trellis Methods with Topographic Information. *Water Resour. Manag.* 27 (8), 2947–2963. <https://doi.org/10.1007/s11269-013-0325-0>.
- Davies, D.L., Bouldin, D.W., 1979. A Cluster Separation Measure. *IEEE Trans. Pattern Anal. Mach. Intell.* PAMI-1 (2), 224–227. <https://doi.org/10.1109/TPAMI.1979.4766909>.
- Di Prinzio, M., Castellarin, A., Toth, E., 2011. Data-driven catchment classification: Application to the pub problem. *Hydrol. Earth Syst. Sci.* 15 (6), 1921–1935. <https://doi.org/10.5194/hess-15-1921-2011>.
- Donnelly, C., Andersson, J.C.M., Arheimer, B., 2016. Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across Europe. *Hydrol. Sci. J.* 61 (2), 255–273. <https://doi.org/10.1080/02626667.2015.1027710>.
- Ercan, K., Mehmet, C., D., Osman, A., B., 2008. HYDROLOGIC HOMOGENEOUS REGIONS USING MONTHLY STREAMFLOW IN TURKEY. *Earth Sci. Res. J.* 12 (2), 181–193.
- Farsadnia, F., Rostami Kamrood, M., Moghaddam Nia, A., Modarres, R., Bray, M.T., 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. <https://doi.org/10.1016/j.jhydrol.2013.11.050>.
- Gaussen, H., Bagnouls, F., 1952. L'indice xérothermique. *Bull. De. l'Assoc. De. Géographes Français* 29 (222), 10–16. <https://doi.org/10.3406/bagf.1952.7361>.
- Herbst, M., Casper, M.C., 2008. Towards model evaluation and identification using Self-Organizing Maps. *Hydrol. Earth Syst. Sci.* 12 (2), 657–667. <https://doi.org/10.5194/hess-12-657-2008>.
- Hosking, J.R.M., Wallis, J.R., 1993. Some statistics useful in regional frequency analysis. *Water Resour. Res.* 29 (2), 271–281. <https://doi.org/10.1029/92WR01980>.
- Hosking, J.R.M., & Wallis, J.R. (2005). *Regional Frequency Analysis: An Approach Based on L-Moments* (Revised ed. edition). Cambridge University Press.
- Jackson, J.E., 2005. *A User's Guide to Principal Components*. John Wiley & Sons.
- Jehn, F.U., Bestian, K., Breuer, L., Kraft, P., Houska, T., 2020. Using hydrological and climatic catchment clusters to explore drivers of catchment behavior. *Hydrol. Earth Syst. Sci.* 24 (3), 1081–1100. <https://doi.org/10.5194/hess-24-1081-2020>.
- Kohonen, T., 2001. *Self-Organizing Maps*, 3rd edition. Springer.
- Kramer, M.A., 1991. Nonlinear principal component analysis using autoassociative neural networks. *AIChE J.* 37 (2), 233–243. <https://doi.org/10.1002/aic.690370209>.
- Ley, R., Casper, M.C., Hellebrand, H., Merz, R., 2011. Catchment classification by runoff behaviour with self-organizing maps (SOM). *Hydrol. Earth Syst. Sci.* 15 (9), 2947–2962. <https://doi.org/10.5194/hess-15-2947-2011>.
- Lin, G.-F., Chen, L.-H., 2006. Identification of homogeneous regions for regional frequency analysis using the self-organizing map. *J. Hydrol.* 324 (1), 1–9. <https://doi.org/10.1016/j.jhydrol.2005.09.009>.
- Lindström, G., Pers, C., Rosberg, J., Strömqvist, J., Arheimer, B., 2010. Development and testing of the HYPE (Hydrological Predictions for the Environment) model - A water quality model for different spatial scales. *Hydrol. Res.* 41 (3–4), 295–319. <https://doi.org/10.2166/nh.2010.007>.
- Lu, B.-W., Pandolfo, L., 2011. Quasi-objective nonlinear principal component analysis. *Neural Netw.* 24 (2), 159–170. <https://doi.org/10.1016/j.neunet.2010.10.001>.
- MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. *Proc. Fifth Berkeley Symp. . Math. Stat. Probab.* 281–298. *Volume 1: Statistics*, 5.1..
- Merz, R., Blöschl, G., 2009. A regional analysis of event runoff coefficients with respect to climate and catchment characteristics in Austria. *Water Resour. Res.* 45 (1). <https://doi.org/10.1029/2008WR007163>.
- Mostofi Zadeh, S. (2012). *Low flow frequency study for Newfoundland and Labrador* [Masters, Memorial University of Newfoundland]. (<https://research.library.mun.ca/10014/>).
- Neira Mendez, F.H. (2006). *Assessment of climate indices in drylands of Colombia*. Universiteit Gent Belgium.
- Neri, M., Coulibaly, P., Toth, E., 2022. Similarity of catchment dynamics based on the interaction between streamflow and forcing time series: Use of a transfer entropy signature. *J. Hydrol.* 614, 128555. <https://doi.org/10.1016/j.jhydrol.2022.128555>.
- Razavi, T., Coulibaly, P., 2013. Classification of Ontario watersheds based on physical attributes and streamflow series. *J. Hydrol.* 493, 81–94. <https://doi.org/10.1016/j.jhydrol.2013.04.013>.
- Razavi, T., Coulibaly, P., 2017. An evaluation of regionalization and watershed classification schemes for continuous daily streamflow prediction in ungauged watersheds. *Can. Water Resour. J. / Rev. Can. Des. Ressour. Hydr.* 42 (1), 2–20. <https://doi.org/10.1080/07011784.2016.1184590>.
- Salami, Y., Buehler, A., 2020. Catchment classification in a transboundary river using runoff and sub-basin characteristics. *Int. J. River Basin Manag.* 18 (1), 117–126. <https://doi.org/10.1080/15715124.2019.1628033>.
- Sawicz, K., Wagener, T., Sivapalan, M., Troch, P.A., Carrillo, G., 2011. Catchment classification: Empirical analysis of hydrologic similarity based on catchment function in the eastern USA. *Hydrol. Earth Syst. Sci.* 15 (9), 2895–2911. <https://doi.org/10.5194/hess-15-2895-2011>.
- Sharghi, E., Nourani, V., Soleimani, S., Sadikoglu, F., 2018. Application of different clustering approaches to hydroclimatological catchment regionalization in mountainous regions, a case study in Utah State. *J. Mt. Sci.* 15 (3), 461–484. <https://doi.org/10.1007/s11629-017-4454-4>.
- Ssegane, H., Tollner, E.W., Mohamoud, Y.M., Rasmussen, T.C., Dowd, J.F., 2012. Advances in variable selection methods II: Effect of variable selection method on classification of hydrologically similar watersheds in three Mid-Atlantic ecoregions. *438–439 J. Hydrol.* 26–38. <https://doi.org/10.1016/j.jhydrol.2012.01.035>.

- Toth, E., 2013. Catchment classification based on characterisation of streamflow and precipitation time series. *Hydrol. Earth Syst. Sci.* 17 (3), 1149–1159. <https://doi.org/10.5194/hess-17-1149-2013>.
- Vesanto, J., Alhoniemi, E., 2000. Clustering of the self-organizing map. *IEEE Trans. Neural Netw.* 11 (3), 586–600. <https://doi.org/10.1109/72.846731>.
- Wagener, T., Sivapalan, M., Troch, P., Woods, R., 2007. Catchment Classification and Hydrologic Similarity. *Geogr. Compass* 1 (4), 901–931. <https://doi.org/10.1111/j.1749-8198.2007.00039.x>.
- Wallis, J.R., Schaefer, M.G., Barker, B.L., Taylor, G.H., 2007. Regional precipitation-frequency analysis and spatial mapping for 24-hour and 2-hour durations for Washington State. *Hydrol. Earth Syst. Sci.* 11 (1), 415–442. <https://doi.org/10.5194/hess-11-415-2007>.
- Wolfe, J.D., Shook, K.R., Spence, C., Whitfield, C.J., 2019. A watershed classification approach that looks beyond hydrology: application to a semi-arid, agricultural region in Canada. *Hydrol. Earth Syst. Sci.* 23 (9), 3945–3967. <https://doi.org/10.5194/hess-23-3945-2019>.
- Zhang, R., Chen, Y., Zhang, X., Ma, Q., Ren, L., 2022. Mapping homogeneous regions for flash floods using machine learning: A case study in Jiangxi province, China. *Int. J. Appl. Earth Obs. Geoinf.* 108, 102717. <https://doi.org/10.1016/j.jag.2022.102717>.