scientific reports

Check for updates

OPEN

# Flow signatures and catchment's attributes for HCA clustering in a hydrologic similarity assessment (Tunisian case)

Rim Chérif[1,2]✉ & Emna Gargouri-Ellouze[1]

**Partitioning methods such as cluster analysis are advantageous in pooling catchments into hydrometric similar regions. They help overcome data shortage in ungauged catchments, which is a common problem in Sud Mediterranean zones. Without accurate forecasts, it is difficult to assess and manage water resources efficiently this situation won't be of any assistance to hydrology decision-makers. This paper illustrates a Tunisian application case, that aims to pool catchments with a hierarchical clustering algorithm (HCA) based on distances calculated in multidimensional physiographical and hydrometric space. The homogeneity of generated clusters is checked by the silhouette index. Then the distances efficiencies are compared. Nineteen semi-arid Tunisian catchments monitored since 1992 are studied. Twelve physiographical attributes, nine rainfall and streamflow signatures are considered in the HCA with two clusters. Correlation distance provides the most homogeneous clusters. Statistically the: percentage of area affected by anti-erosive practices, percentage of forest cover and catchment area are the most discriminating attributes. However, hydrometrical signatures appear to be irrelevant. These partitions highlight two different hydrological behaviors that must support forecasting. Results are promising in the Sud-Mediterranean case, where the shortage of hydrometrical data is an ongoing problem. They have the advantage of enabling hydrologic forecasting without requiring heavy information.**

Water resources management (ex: land use planning, irrigation, hydraulic structure design, flood forecasting) requires knowledge of water quantity at a target site or catchment. Nevertheless, several catchments in many parts of the world are ungauged or poorly gauged, this lack of data often increases with decreasing catchment sizes that leads to great difficulties in their management[1,2]. Therefore, runoff prediction at an ungauged river or catchment is carried out through some kind of extrapolation from a gauged site to an ungauged site, and this is not straightforward. This is the whole raison d'être of the Prediction of Ungauged Basin (PUB) initiative[2]. PUB was designed to develop a better scientific basis for hydrology with greater consistency, increasing the prospects for scientific breakthroughs and reducing uncertainties[3].

Regionalisation techniques are PUB tools that are necessary for transferring information. They belong to two categories; statistical or process based. The transfer of information from one or several gauged catchments (donors) to another ungauged catchment (receiver)[4] requires the identification of similar gauged catchments, which can be selected through:

- Geographical or spatial proximity.
- Similarities in their hydrologic and/or physiographical and climatic attributes applied with clustering approaches. Thus, Metric distances are commonly identified between catchments in multidimensional attribute space to assess their proximity[5,6].

In practice, hydrologists explored a large range of approaches for regionalisation over time, as there are no established criteria by which the superiority of any approach can be clearly brought out[7,8].

[1]LR99ES19 Laboratory of Modelling in Hydraulics and Environment (LMHE), National Engineering School of TUNIS (ENIT), University of Tunis El Manar, BP 37, 1002 Tunis, Tunisia. [2]High Institute of Environmental Sciences and Technology (HIEST), Borj Cedria. Carthage University, Tunis, Tunisia. ✉email: rim.cherif2009@gmail.com

Burn and Goel[9] adopted clustering as a starting point for catchment partitions based on physiographical catchment characteristics with weighted Euclidian distance. Then a regional revision heuristic process was proposed to increase the region's homogeneity[10]. Lately, Jared et al.[11] studied, in a classification framework, small Canadian catchments within the Prairie based on climatic and biophysical attributes. They identified similar regions with the agglomerative hierarchical clustering of principal components (HCPC) method. Thus, it can be underlined that regionalisation studies frequently require catchment classification, on which their accuracy closely depends.

Despite the large efforts in PUB, there is still a long way to go in terms of achieving robust and reliable predictions. Ungauged basins have seen less success thus far than gauged basins, which is detrimental to developing countries where the management of sustainable water resources and the development of effective flood and drought mitigation strategies will continue to be hampered by our inability to accurately predict the future[3].

Unsupervised classification is a data mining technique that is undoubtedly a challenging research area. It could be defined as the organization of a collection of patterns into groups based on similarity analysis[12,13]. Many hydrologic scholars applied this class of clustering algorithms for the purpose of analyzing catchments similarity based on their physiographical, climatic, stream-flow signatures, etc.[14]. Goval and Gupta[15] divide clustering methods into hierarchical (agglomerative and divisive) and partitional (hard clustering [ex: k-mean] and soft clustering [ex: fuzzy C-mean]).

Partitional clustering methods divide a data set of objects based on their similarity. For K-Means clustering[16], the number of clusters (K) is defined previously; the initial clusters are first randomly selected, then modified to generate new clusters that minimize the variance within each cluster. Each object can belong to several clusters in the case of soft classification.

Hierarchical cluster analysis (HCA) algorithms pool similar objects into a hierarchy of clusters. They offer a series of interlocked partitions in the form of trees called dendrograms. The main advantage of HCA compared with partitional clustering methods lies in the dendrogram representation, which highlights additional information, such as the increase in dispersion in a cluster generated by an aggregation. It also does not require determining the number of clusters in advance. Indeed, by observing the dendrogram and playing with the depth of the tree, we can explore different possibilities and choose the number of clusters that suit our application case best. Thus, it is conceptually simple, good for small data sets, and less sensitive to noise in the data set[17].

In our case, HCA is better suited to identify catchment clusters with similar hydrologic behaviors. Metrics (or distances) are used to measure this similarity[9]. Hence, distances evaluate the proximity, or relevance, of each gauged catchment to the target location and identify the most hydrologically similar one[18]. Since it aims to reduce the variance between entities within a cluster, we use it in conjunction with Ward's linkage method[19].

Many useful distances, such as Euclidean, squared Euclidean, Manhattan, Chebyshev, cosine, Canberra, Minkowski, and Mahalanobis, were cited in the literature. Nathan and Mc Mahon[20] compared combinations of similarity measures (Euclidean, squared Euclidean, Manhattan, Chebyshev, and Cosine) and linkage methods (simple, complete, average, and Ward) to identify homogeneous sub-regions from 184 catchments in southeast Australia and forecast low flow characteristics. They found that the best combinations are Ward with squared Euclidean and average with cosine. Later, Cunderlik and Burn[21] recommended using Mahalanobi's distance since it considers the variance and covariance of variables, which is not obvious with other distances. Shirkhorshidi et al.[22] compared similarity and dissimilarity measures in clustering various continuous data sets. They employed the Minkowski family, including Euclidean and Manhattan distances and the modified versions of Euclidean distance: average, weighted Euclidean, and chord distance; cosine similarity measure; and Pearson correlation. They concluded that average distance was among the topmost accurate measures for all clustering algorithms.

In south Mediterranean regions, study cases are not so large. Singla et al.[23] studied the hydrological regimes of 27 river basins in Morocco to assess the impact of climate change on water resources. They applied the regional vector method to outline homogeneous rainfall variability and assess the representativeness and persistence of regional signals. They outline that in the Rif and the Mediterranean Sea, rainfall revealed a trend towards a relative increase since 1980 but a significant decrease in other regions. Monthly and annual discharge analyses showed a decrease since the late 1970s. In 2017, Totz et al.[24] developed a new cluster-based empirical forecast method (HCA) to predict precipitation anomalies in winter. This method outperformed both statistical and dynamical models over comparable historical periods in the European and Mediterranean regions.

In the south Mediterranean region, Ahattab et al.[25] utilized morphological parameters and the series of monthly precipitation recorded at 23 rainfall stations (with a common observation period of 15 years) spread across the Tensift watershed (Morocco) to identify four homogeneous clusters that can be considered to exhibit hydrologically similar behaviors and for which the same models for estimating flood peaks can be applied. In Tunisia, few hydrological regionalization studies involving catchment classification were done. Bargaoui et al.[26] applied the ISODATA method to regionalize 39 Tunisian catchments and assess the centennial flood. The copula model classification based on physiographic and geographic catchment characteristics was later investigated by Gargouri and Bargaoui[27] to delineate 22 Tunisian catchments in hydro-physiographical regions. They noted that the catchments in the same region are not necessarily geographically contiguous.

Subsequently, At 55 stations of the Tunisian gauging network, Bargaoui and Chebchoub[28] applied the multifractal analysis of maximum annual flood discharges. They identified a random cascade model after successfully connecting the various statistical moments of the basins' surface discharges through a scale-invariant law.

Next, Cherif and Bargaoui[29] used HCA to construct a mean regional frequency curve for annual maximum runoffs and applied topographic descriptors for cluster analysis. They utilized Trellis and hierarchical classifications for partitioning with a sample of 40 Tunisian watersheds. Various multidimensional spaces were studied with pairs or triplets of attributes to construct the distance measures. Finally resulting clusters were

checked for hydrological homogeneity applying the Hosking and Wallis test. They concluded that global slope index is highlighted as scale factor for flood index.

Then, Cherif and Gargouri[30] studied the hydrologic behavior of twenty catchments situated on Tunisian ridge. To define hydrologic regions, they used the Hosking and Wallis test and the moving average clustering method related to catchment hydro-geomorphologic attributes[34]. Next, they hold on to regional frequency curves of the maximum specific discharge index.

Later, Kotti et al.[31] used the regional vector method to divide the study area into six climatically homogeneous subregions after identifying the regional components of the variability of river flows in the Medjerda watershed (the largest river basin in Tunisia). Then they developed regionalized regression models to determine the runoff coefficient and studied the inter-correlations between stations to fill in a series of flow data.

They validated the possibility of estimating runoff at a station based on the maximum rate and the rain from the same station and hydrologic parameters from a neighboring gauging station, with a noticeable improvement in runoff depth values compared to the literature. Recently, Gargouri et al. [32] used Ward's algorithm with Euclidian distance and agglomerative hierarchical clustering to study 22 Tunisian catchments, where the dissimilarity between clusters is calculated in the multidimensional space of geomorphological and physiographical variables. Then, regions homogeneity and consistency are measured by the silhouette index. This study led to three homogeneous regions, performed using a multivariate copula.

Although there has been significant research in PUB applications on Mediterranean cases clustering Tunisian catchments and Sud Mediterranean regions, HCA techniques applications remain limited. More efforts are needed in clustering analysis applications that are still uncommon and underrepresented due to the difficult gauging circumstances and lack of hydrometric data in the region.

Indeed, better understanding of Tunisian catchment's behavior can be highly valuable for hydrologists in Tunisia, It contributes to the advancement of hydrological modeling, supports decision making processes in water resources management and offers beneficial insights into the hydrological characteristics of other Mediterranean regions, especially the Sud Mediterranean that have climate and agricultural practices similarities.

This study aims to analyze hydrologic similarity between Tunisian ridge catchments based on the HCA algorithm and the homogeneity index of delineated clusters. Several metric distances were applied in the linkage method, and their efficiencies are compared.

(i) To attempt this objective, the following steps will be carried out: applying HCA for Tunisian catchments with similarity distances based on their geo-morphological attributes and hydrometrical signatures.
(ii) Integrating Silhouette index to validate the homogeneity of clusters. Hence, we compare efficiency of all distances to predict the most accurate one.
(iii) Analyzing results to better understand Tunisian catchment's behaviors.

## Materials and methods

**Clustering approach.** HCA is an unsupervised multivariate analysis that classifies the given data into similar, overlapping, or non-overlapping clusters. It has large applications for finding homogeneous clusters of objects based on metric distances between objects. HCA seeks to build a hierarchy of clusters that can be agglomerated or divisive. Agglomerating algorithms merge clusters. On the contrary, divisive algorithms split clusters. Both can be illustrated as a nested sequence or tree diagram, called a dendrogram. It shows the linkage points and clusters that are connected at increasing levels of dissimilarity. The heights of the branch points indicate how similar or different they are from each other; the greater the height, the greater the difference.

In the current study, the HCA algorithm is applied to delineate clusters of similar catchments; we focus on defining the most homogeneous clusters. Homogeneity is defined by the similar hydrodynamic behavior of catchments. Hence, we are seeking the more suitable distance that gives the best similarity in the clusters.

As a first step, correlations are calculated between all attributes and signatures after their standardization. Then, all specified attributes and signatures are implemented in HCA. Cluster homogeneity is assumed to be ensured; afterward, to validate this hypothesis, the silhouette index is calculated. Each cluster is characterized by its silhouette index, which compares its tightness and separation. It illustrates which feature vectors belong to the cluster and which ones are just in between clusters. Cluster's silhouette indexes show consistency within clusters and provide a means of assessing cluster quality[32]. They are calculated for each cluster and then compared between all applied similarity distances to outline the best one for the hierarchical clustering approach. The steps of the clustering methodology applied in our current study are summarized in Fig. 1.

**Distance's equations.** Let's consider a matrix $X$ of size $n \times p$: rows are the individuals ($n$), and columns are the variables ($p$) (Eq. 1.)

$$X = \begin{bmatrix} x_1^1 & \cdots & x_1^p \\ \vdots & x_i^j & \vdots \\ x_n^1 & \cdots & x_n^p \end{bmatrix} \quad (1)$$

With $x_i$: ith row and $x_j$: jth column.

The distance $d(x_i, x_j)$ is defined between two vectors $x_i$ and $x_j$ $(i, j = 1 \ldots n)$ in the $p$-dimensional space $\mathbb{R}^p$.The distances utilized in this work, for hierarchical analysis, are illustrated in Table 1.

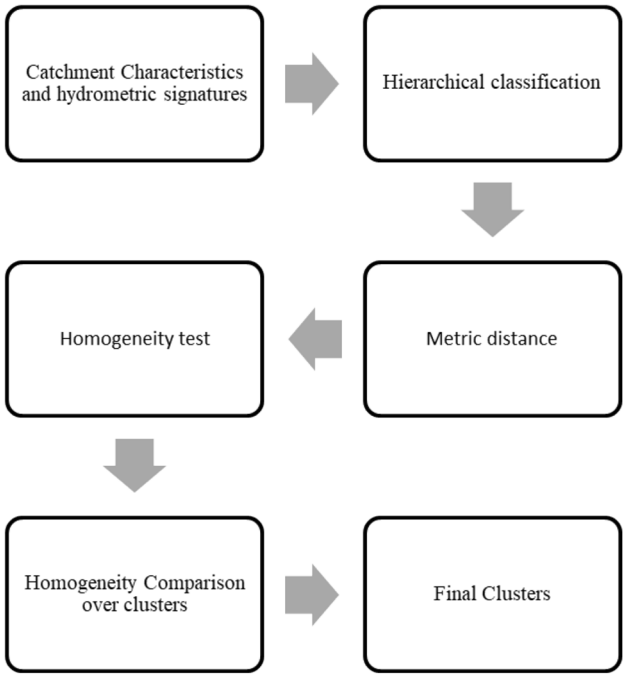



**Figure 1.** Illustration of the clustering methodology steps.

| Distance | Expression $d(x_i, x_j)$ | Comments |
|---|---|---|
| Euclidean | $\sqrt{(x_i - x_j)(x_i - x\_j)^t}$ | $x^t$: *transpose vector* |
| Standardized euclidean | $\sqrt{(x_i - x_j)V^{-1}(x_i - x_j)^t}$ | V is the *p-by-p* diagonal matrix whose $j_{th}$ diagonal element is squared standard deviation; |
| Chebyshev | $max_k\left\{\left\|x_i^k - x_j^k\right\|\right\}$ | $k = 1, \ldots\ldots\ldots\ldots, p$ |
| Cosine | $1 - \frac{x_i x_j^t}{\sqrt{(x_i x_i^t)(x_j x_j^t)}}$ | $x^t$: transpose vector |
| Correlation | $1 - \frac{(x_i - \overline{x_i})(x_j - \overline{x_j})^t}{\sqrt{(x_i - \overline{x_i})(x_i - \overline{x_i})^t}\sqrt{(x_j - \overline{x_j})(x_j - \overline{x_j})^t}}$ | – |
| Hamming | $\left(\#\left(x_i^k \neq x_j^k\right)/p\right)$ | $k = 1, \ldots\ldots\ldots\ldots, p$ |
| Jaccard | $\frac{\left[\left(x_i^k \neq x_j^k\right) \cap \left(\left(x_i^k \neq 0\right) \cup \left(x_j^k \neq 0\right)\right)\right]}{\left[\left(x_i^k \neq 0\right) \cup \left(x_j^k \neq 0\right)\right]}$ | $k = 1, \ldots\ldots\ldots\ldots, p$ |
| Spearman | $1 - \frac{(r_i - \overline{r_i})(r_j - \overline{r_j})^t}{\sqrt{(r_i - \overline{r_i})(r_i - \overline{r_i})^t}\sqrt{(r_j - \overline{r_j})(r_j - \overline{r_j})^t}}$ | $r_i$ and $r_j$ are the coordinate-wise rank vectors of $x_i$ and $x_j$ |

**Table 1.** Distance equations utilized for hierarchical analysis.

**Performance criteria (Silhouette index) and cluster's homogeneity validation.** One of current study objectives is to assess and compare the efficiency of metric distances in the HCA approach. It is crucial to confirm that delineated clusters, really reflect hydrological homogeneity.

The Silhouette index describes each cluster by contrasting its tightness and separation to assess homogeneity. It illustrates which feature vectors belong to their cluster, and which ones are just in between clusters. Cluster's Silhouettes are plotted in a chart showing consistency within clusters and providing assessing cluster quality[32].

For each feature vector $x_i$, the corresponding Silhouette index *s(i)* is defined as:

$$s(i) = \frac{b(i) - a(i)}{max[a(i), b(i)]} \tag{2}$$

where, for a given $x_i$ belonging to cluster A (with ) and a distance *d* (.,.),

$$a(i) = \frac{1}{Card(A) - 1} \sum_{\substack{x_j \in A \\ i \neq j}} d(x_i, x_j) \tag{3}$$

$$b(i) = min_{A \neq C} \frac{1}{Card(C)} \sum_{\substack{x_k \in C \\ i \neq k}} d(x_i, x_k) \tag{4}$$

where  is the cardinal number. It is thought as an equivalence class of sets. $a(i)$ is the average distance from the $i$th feature vector to all other feature vectors in the cluster $A$; $b(i)$ is the minimum average distance from the $i$th feature vector to all the feature vectors in another cluster $C$. From this equation it follows that $-1 \leq s(i) \leq 1$. If $s(i)$ is large thus the $i$th feature vector is well assigned to the cluster. On the other hand, when $s(i)$ is close to -1 the $i$th feature vector is not well classified.

The closer Silhouette index approaches to 1, the better cohesion and separation are[33].

Therefore, it is applied in this study to evaluate and compare clustering approaches, Silhouette indexes are calculated for each catchment in the cluster, and then their averages are deduced. A positive value reveals that the catchment is well matched to its cluster. A negative one means that the catchment is not in the right cluster, so it could be moved to the more closely related one[34].

Hence a high Silhouette index demonstrates that the classified feature vector (catchment) is well pooled and poorly matched to neighboring clusters. If the Silhouette index value is close to (-1), it means that the individual is not in the right cluster[39].

## Study case

In current research we compared the use of hierarchical classification with several distances specified in Table 1 (see § 2.2) to delineate catchments into hydrological regions. Silhouette indexes are then calculated for each catchment in each delineated cluster to define the best distance giving the higher homogeneity for clusters (regions). Hence, MATLAB software package is utilized.

We considered nineteen 19 catchments situated in the Tunisian ridge and monitored since 1992, controlled by headwater dams. Latitudes vary between 35°N and 37°N; longitudes from 8°E and 11°E, areas range from 1 $km^2$ to 10 $km^2$ and annual average rainfall vary between 280 and 500 mm, these catchments are in a semi-arid zone. These catchments are little permeable to impermeable and have fairly high too high reliefs that promote rapid runoff. The rain gauge network is composed of 19 gauges Fig. 2, located at each headwater dam[35].

Two data sets (Catchment's attributes and streamflow signatures) are utilized, with HCA algorithm, to delineate homogeneous regions. The first set illustrates physiographical catchment's attributes and are selected because they can predetermine hydrological behavior [37; 5; 38], it is hold in Table A.1 of the appendices and is composed of: Latitude (LatN); longitude (LongE); area (A); Perimeter (P); specific denivelation ($D_S$); global slope index ($I_s$); Gravellus Index ($I_G$); the percentage of path (Pp); the percentage of forest cover (Pf); the percentage of cereal culture area (Pc); the percentage of arboriculture area (Pa); the percentage of area affected by anti-erosive practices (Aae)).

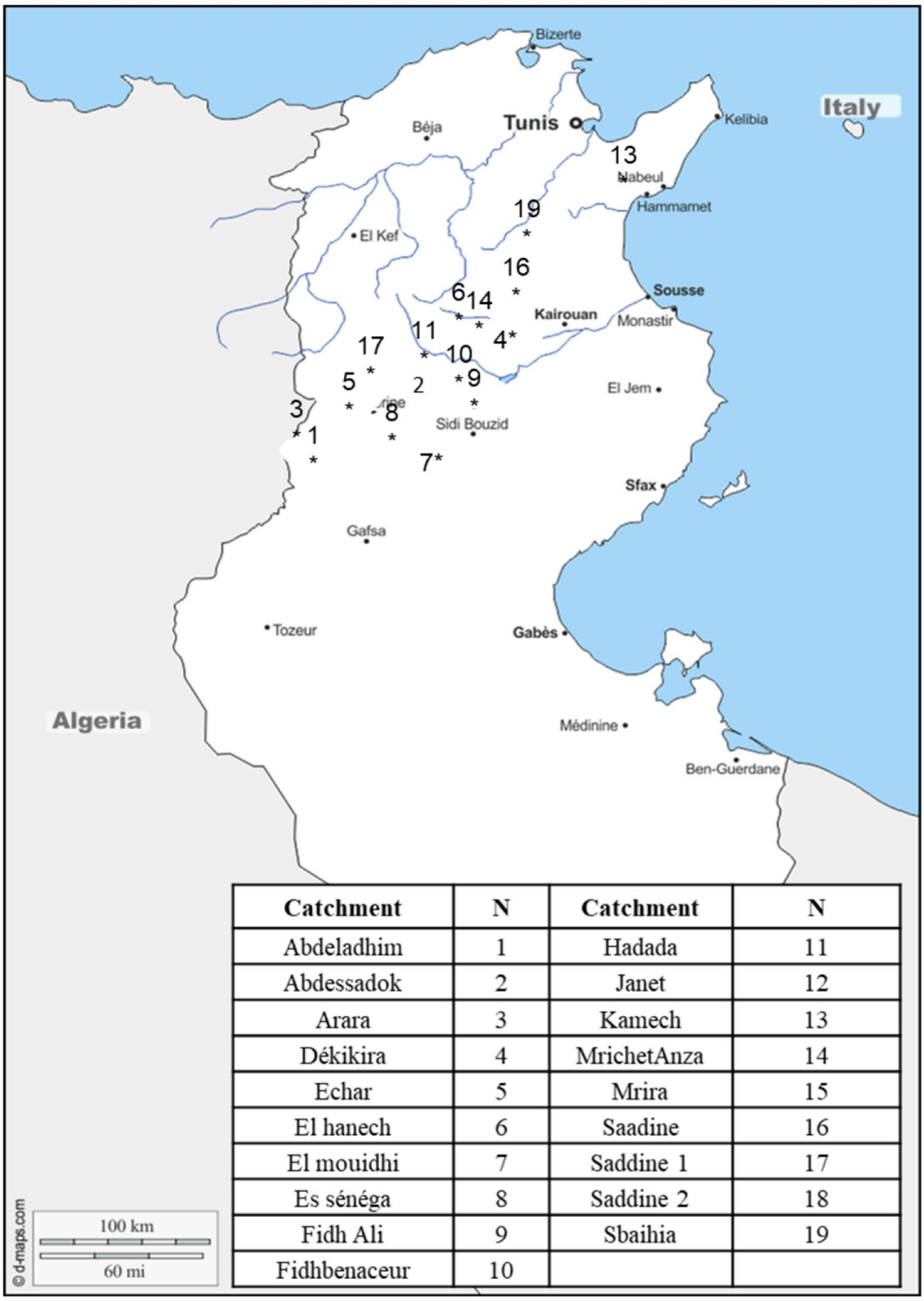
The second set is hold in the Table A.2 of the appendices and summarizes the hydrometrical signatures defined as:

- maximum rainfall intensity ($I_{max}$).
- rainfall duration (D), runoff depth.
- Runoff depth (R),: runoff volume from a drainage basin, divided by its area, in a specified time expressed in mm.
- Hydrograph time to peak (tp): the increase time of hydrograph.
- Hydrograph base time (tb): time between the begin and the end of the hydrograph.
- Infiltration index (ϕ): average rate of infiltration derived from a time intensity graph of rainfall in such a manner that the volume of rainfall in exceedance of this rate will equal the volume of storm runoff[38].
- Runoff coefficient (Cr): ratio of runoff depth to precipitation depth.
- Average discharge ($Q_{mean}$): average daily runoff.
- Specific Maximum discharge ($QS_{max}$): maximum discharge divided by the catchment area.

These signatures quantify the hydrologic response and provide insight into the functional behavior of the catchment[37]. They are included to support the hypothesis of hydrological homogeneity. Tables A.1 and A.2 holds also, specific statistics such as mean values, standard deviation, minimum values, and maximum values, which are denoted as Min, Max, μ and σ respectively. All data come from hydrological reports of the Tunisian Water Resources Division (DGRE).

## Results and discussion

As a first step, correlations are calculated between all attributes and hydrometrical signatures after their standardization (Table 2a, b). Ranging between − 0.7 and 1, they reveal that geo-morphological attributes are closely to slightly connected with flow signatures and rainfall descriptors, with inter-correlations varying between − 0.5 and 0.8. Hence, watersheds have hydrological behaviors influenced by their geomorphology. This result is in accordance with the work of Kotti et al.[35] in their study of the Medjerdah watershed, where they deduce that the flows are a relative response to different factors (watershed size, relief, geology, soils, and vegetation cover).




| Catchment | N | Catchment | N |
|---|---|---|---|
| Abdeladhim | 1 | Hadada | 11 |
| Abdessadok | 2 | Janet | 12 |
| Arara | 3 | Kamech | 13 |
| Dékikira | 4 | MrichetAnza | 14 |
| Echar | 5 | Mrira | 15 |
| El hanech | 6 | Saadine | 16 |
| El mouidhi | 7 | Saddine 1 | 17 |
| Es sénéga | 8 | Saddine 2 | 18 |
| Fidh Ali | 9 | Sbaihia | 19 |
| Fidhbenaceur | 10 | | |

**Figure 2.** Tunisian Hydrometrical network considered in this study (map from https://d-maps.com, opensource, and modified by QGIS: GNU General Public License, version 2.0).

Next, HCA is applied with distances calculated from geomorphologic attributes and hydrometrical signatures (previously defined) to delineate clusters of catchments with similar behavior (homogeneous clusters). Hence,

a

| | LatN | LongE | A | P | $D_s$ | $I_s$ | $I_G$ | Pp | Pf | Pc | Pa | Aae |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| latN | 1.0 | 0.8 | -0.5 | -0.4 | -0.1 | 0.1 | -0.1 | -0.3 | 0.0 | 0.2 | 0.4 | 0.0 |
| longE | 0.8 | 1.0 | -0.7 | -0.7 | -0.4 | 0.0 | -0.4 | 0.1 | -0.1 | 0.0 | 0.3 | 0.1 |
| A | -0.5 | -0.7 | 1.0 | 0.9 | 0.2 | -0.3 | 0.5 | -0.3 | 0.3 | -0.1 | -0.2 | -0.4 |
| P | -0.4 | -0.7 | 0.9 | 1.0 | 0.3 | -0.2 | 0.8 | -0.2 | 0.3 | -0.1 | -0.2 | -0.4 |
| $D_s$ | -0.1 | -0.4 | 0.2 | 0.3 | 1.0 | 0.9 | 0.2 | -0.1 | 0.1 | 0.0 | -0.4 | 0.1 |
| $I_s$ | 0.1 | 0.0 | -0.3 | -0.2 | 0.9 | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 | -0.4 | 0.3 |
| $I_G$ | -0.1 | -0.4 | 0.5 | 0.8 | 0.2 | 0.0 | 1.0 | -0.1 | 0.3 | -0.1 | -0.2 | -0.2 |
| Pp | -0.3 | 0.1 | -0.3 | -0.2 | -0.1 | 0.0 | -0.1 | 1.0 | -0.5 | -0.5 | 0.1 | -0.1 |
| Pf | 0.0 | -0.1 | 0.3 | 0.3 | 0.1 | 0.0 | 0.3 | -0.5 | 1.0 | 0.0 | -0.3 | 0.0 |
| Pc | 0.2 | 0.0 | -0.1 | -0.1 | 0.0 | 0.0 | -0.1 | -0.5 | 0.0 | 1.0 | 0.3 | 0.2 |
| Pa | 0.4 | 0.3 | -0.2 | -0.2 | -0.4 | -0.4 | -0.2 | 0.1 | -0.3 | 0.3 | 1.0 | -0.3 |
| Aae | 0.0 | 0.1 | -0.4 | -0.4 | 0.1 | 0.3 | -0.2 | -0.1 | 0.0 | 0.2 | -0.3 | 1.0 |
| $I_{max}$ | 0.3 | 0.3 | -0.2 | -0.1 | 0.1 | 0.1 | -0.1 | 0.1 | -0.2 | -0.1 | 0.2 | 0.2 |
| D | 0.4 | 0.5 | -0.3 | -0.3 | 0.0 | 0.1 | -0.1 | 0.0 | -0.2 | 0.0 | 0.4 | 0.2 |
| R | 0.0 | 0.2 | -0.3 | -0.2 | 0.1 | 0.2 | 0.0 | 0.0 | 0.1 | -0.3 | 0.0 | 0.2 |
| tp | 0.5 | 0.4 | -0.3 | -0.2 | -0.2 | -0.1 | 0.0 | -0.1 | -0.2 | 0.5 | 0.5 | 0.1 |
| tb | 0.5 | 0.2 | -0.1 | 0.1 | 0.2 | 0.1 | 0.3 | -0.3 | 0.1 | 0.5 | 0.2 | 0.1 |
| Φ | -0.1 | 0.2 | -0.2 | -0.2 | -0.1 | 0.0 | -0.1 | 0.3 | -0.2 | -0.2 | -0.1 | 0.3 |
| Cr | 0.0 | 0.1 | -0.4 | -0.4 | -0.1 | 0.1 | -0.3 | 0.2 | -0.1 | -0.2 | 0.0 | 0.0 |
| $Q_{mean}$ | -0.5 | -0.5 | 0.6 | 0.5 | 0.2 | -0.1 | 0.3 | 0.1 | 0.1 | -0.4 | -0.1 | -0.1 |
| $QS_{max}$ | -0.1 | 0.0 | -0.2 | -0.2 | 0.3 | 0.4 | 0.0 | -0.1 | 0.0 | -0.2 | -0.3 | 0.6 |

b

| | $I_{max}$ | D | R | tp | tb | Φ | Cr | $Q_{mean}$ | $QS_{max}$ |
|---|---|---|---|---|---|---|---|---|---|
| latN | 0.3 | 0.4 | 0.0 | 0.5 | 0.5 | -0.1 | 0.0 | -0.5 | -0.1 |
| longE | 0.3 | 0.5 | 0.2 | 0.4 | 0.2 | 0.2 | 0.1 | -0.5 | 0.0 |
| A | -0.2 | -0.3 | -0.3 | -0.3 | -0.1 | -0.2 | -0.4 | 0.6 | -0.2 |
| P | -0.1 | -0.3 | -0.2 | -0.2 | 0.1 | -0.2 | -0.4 | 0.5 | -0.2 |
| $D_s$ | 0.1 | 0.0 | 0.1 | -0.2 | 0.2 | -0.1 | -0.1 | 0.2 | 0.3 |
| $I_s$ | 0.1 | 0.1 | 0.2 | -0.1 | 0.1 | 0.0 | 0.1 | -0.1 | 0.4 |
| $I_G$ | -0.1 | -0.1 | 0.0 | 0.0 | 0.3 | -0.1 | -0.3 | 0.3 | 0.0 |
| Pp | 0.1 | 0.0 | 0.0 | -0.1 | -0.3 | 0.3 | 0.2 | 0.1 | -0.1 |
| Pf | -0.2 | -0.2 | 0.1 | -0.2 | 0.1 | -0.2 | -0.1 | 0.1 | 0.0 |
| Pc | -0.1 | 0.0 | -0.3 | 0.5 | 0.5 | -0.2 | -0.2 | -0.4 | -0.2 |
| Pa | 0.2 | 0.4 | 0.0 | 0.5 | 0.2 | -0.1 | 0.0 | -0.1 | -0.3 |
| Aae | 0.2 | 0.2 | 0.2 | 0.1 | 0.1 | 0.3 | 0.0 | -0.1 | 0.6 |
| $I_{max}$ | 1.0 | 0.8 | 0.0 | 0.5 | 0.1 | 0.8 | -0.4 | -0.1 | 0.2 |
| D | 0.8 | 1.0 | 0.4 | 0.6 | 0.4 | 0.5 | -0.2 | -0.1 | 0.3 |
| R | 0.0 | 0.4 | 1.0 | 0.1 | 0.2 | 0.1 | 0.3 | 0.1 | 0.8 |
| tp | 0.5 | 0.6 | 0.1 | 1.0 | 0.7 | 0.3 | -0.2 | -0.2 | 0.2 |
| tb | 0.1 | 0.4 | 0.2 | 0.7 | 1.0 | -0.1 | -0.2 | -0.2 | 0.2 |
| Φ | 0.8 | 0.5 | 0.1 | 0.3 | -0.1 | 1.0 | -0.3 | 0.1 | 0.4 |
| Cr | -0.4 | -0.2 | 0.3 | -0.2 | -0.2 | -0.3 | 1.0 | -0.1 | 0.2 |
| $Q_{mean}$ | -0.1 | -0.1 | 0.1 | -0.2 | -0.2 | 0.1 | -0.1 | 1.0 | 0.3 |
| $QS_{max}$ | 0.2 | 0.3 | 0.8 | 0.2 | 0.2 | 0.4 | 0.2 | 0.3 | 1.0 |

**Table 2.** Correlations between attributes and hydrometrical signatures.

we search to outline the best distance involving the most homogeneous clusters. To attempt this objective, all distance similarities previously cited in Table 1 are applied to catchment attributes and signatures.

Due to the limited total number of catchments. They are divided into two clusters with dendrogram agglomeration, and clusters are concatenated graphically. The level of cluster homogeneity is then determined by computing silhouette indices.

Clustering results are summarized in Table 3. This table reveals that catchments 7, 8, and 9 are consistently in the same pool, highlighting a persistent similarity regardless of the distance. Average silhouette values and catchment partitioning for each distance are displayed in Table 4, which indicates that city-block, Hamming, Spearman, and Jaccard distances provide an equal distribution of catchments.

All distances indicate positive average Silhouette indexes (ASI) values for both clusters, ranging from 0.04 to 0.418 for the first cluster and from 0.001 to 0.188 for the second one. So, catchments in first cluster indicate a greatest consistency (homogeneity).

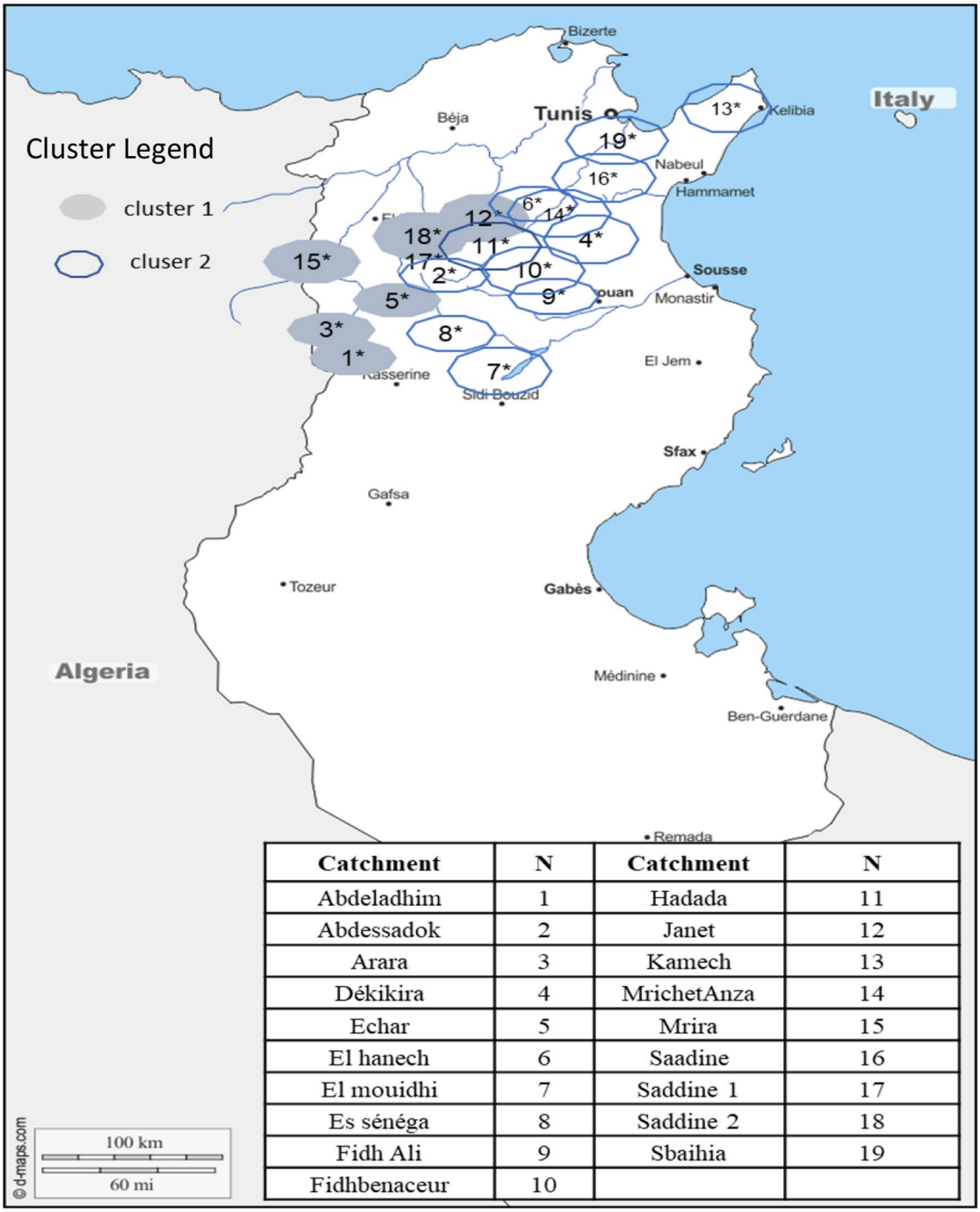
Hence, we deduce that Correlation distance provides the best consistent groups with ASI values of 0.42 and 0.18. The first cluster is composed of 32% of total catchments when the second one implies 68% of them. It is followed by Cosine and Spearman distances with Silhouette indexes respectively of [0.28; 0.21] and [0.27; 0.17]. Euclidean and Seuclidean reveals similar results. Cityblock, Hamming and Jaccard distances produce a nearly equal distribution (with 9 and 10 catchments in each cluster). However, Hamming and Jaccard distances display clusters with the lowest similarities for which Silhouette indexes are equal respectively to [0.048 and 0.0005]. We conclude that with these distances, catchments of the first cluster are more hydrometrically similar.

| | | Distances | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Euclidean | Seuclidean | Cityblock | Cheybychev | Cosine | Correlation | Hamming | Spearman | Jaccard |
| Catchment N | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 |
| | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 1 | 2 |
| | 3 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 2 | 1 |
| | 4 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 1 | 2 |
| | 5 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 2 | 1 |
| | 6 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 |
| | 7 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 1 |
| | 8 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 1 |
| | 9 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 1 |
| | 10 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 1 | 2 |
| | 11 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 |
| | 12 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 2 | 1 |
| | 13 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 1 | 2 |
| | 14 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 2 |
| | 15 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 |
| | 16 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 2 |
| | 17 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | 2 |
| | 18 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 2 | 1 |
| | 19 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |

**Table 3.** Clustering results of all metric distances (watershed membership to each cluster).

| | Cluster 1 | | Cluster 2 | |
|---|---|---|---|---|
| | Catchment number | Average silhouette index | Catchment number | Average silhouette index |
| Euclidean | 7 | 0.2309 | 12 | 0.0738 |
| Seuclidean | 7 | 0.2309 | 12 | 0.0738 |
| Cityblock | 9 | 0.1226 | 10 | 0.1465 |
| Cheybychev | 2 | − 0.0229 | 17 | 0.1122 |
| Cosine | 8 | 0.2838 | 11 | 0.2145* |
| Correlation | 6 | 0.4186* | 13 | 0.1778 |
| Hamming | 9 | 0.0475 | 10 | 0.0005 |
| Spearman | 9 | 0.2685 | 10 | 0.1661 |
| Jaccard | 9 | 0.0475 | 10 | 0.0005 |

**Table 4.** Total number of catchments in each cluster and average silhouette index. *maximum value.




| Catchment | N | Catchment | N |
|---|---|---|---|
| Abdeladhim | 1 | Hadada | 11 |
| Abdessadok | 2 | Janet | 12 |
| Arara | 3 | Kamech | 13 |
| Dékikira | 4 | MrichetAnza | 14 |
| Echar | 5 | Mrira | 15 |
| El hanech | 6 | Saadine | 16 |
| El mouidhi | 7 | Saddine 1 | 17 |
| Es sénéga | 8 | Saddine 2 | 18 |
| Fidh Ali | 9 | Sbaihia | 19 |
| Fidhbenaceur | 10 | | |

**Figure 3.** The Clusters achieved with the correlation distance (map from https://d-maps.com, opensource, and modified by QGIS: GNU General Public License, version 2.0).

Correlation distance reveals the most homogeneous clusters. Figure 3 holds on catchments belonging to each cluster. It is worth noting that the catchments in a same cluster are not necessarily geographically contiguous; in effect, the geographical proximity of the catchments is not a guarantee of their hydrological similarity[39]. This result is in accordance with the one described by Gargouri-Ellouze and Bargaoui[35].

Therefore, we sign that the distance selection could improve accuracy of the clustering method and the hydrological homogeneity in the clusters: This outcome must be considered when dealing with regionalization studies in south Mediterranean regions.

This effect is in accordance with Totz et al.[24] study, in which they developed a new cluster-based empirical forecast method (HCA), to predict winter precipitation's anomalies in European and Mediterranean regions.

Their method achieves a higher skill than other empirical methods used in the past such as the multi-regression model developed by Eden et al.[40] or the CCA-based algorithm applied by Barnston et al.[41].

Catchments within each cluster, statistics of physiographical attributes and hydrometrical signatures are summarized below (Table 5), as well as the ratios of means (ρ) in each cluster for all attribute and signature. $\rho = \mu_1/\mu_2$ where $\mu_1$ and $\mu_2$ are means of (attribute or signature) in respectively the first and the second cluster.

We notice that the first cluster contains larger averages runoffs and areas., indicating wetter and larger catchments. With mean runoff rates twice as high as the second cluster (ρ = 1.8). The Percentage of forest data (Pf) are three times higher as those recorded in the second pool revealing that it greatly controls clustering results. Hence large catchments with important forest covers have similar runoffs indicating similar hydrological behaviors. This outcome can be valuable to neighboring countries within the same climate (Mediterranean regions) and then extended to other anthropogenic indexes.

Figure 4 holds on log absolute values of these ratios which can enlighten us about the most discriminant attributes and signatures. Zero indicates similar averages between the two clusters and the further from zero, the more discriminating the attribute is (resp. signature). It demonstrates three discriminant attributes: the percentage of area affected by anti-erosive practices (Aae) which is the most significant, followed by the percentage of forest

| | | Cluster 1 | | | | Cluster 2 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Min | Max | μ | σ | Min | Max | μ | Σ | ρ |
| Physiographical attributes | A ($km^2$) | 5.2 | 9.2 | 6.8 | 1.3 | 1.6 | 4.7 | 3.1 | 0.9 | 2.2 |
| | P (m) | 11.6 | 16.8 | 13.8 | 2.0 | 5.5 | 9.9 | 8.0 | 1.4 | 1.7 |
| | $D_S$ (m) | 86.7 | 207.5 | 135.9 | 43.3 | 62.6 | 224.3 | 133.6 | 63.3 | 1.0 |
| | $I_s$ (m/km) | 35 | 78 | 53.0 | 17.7 | 37 | 128 | 74.9 | 30.2 | 0.7 |
| | $I_G$ | 1.3 | 1.8 | 1.5 | 0.2 | 1.2 | 1.4 | 1.3 | 0.1 | 1.2 |
| | Pp (%) | 0 | 39 | 19.7 | 13.4 | 0 | 84 | 34.5 | 25.7 | 0.6 |
| | Pf (%) | 0 | 50 | 24 | 21.9 | 0 | 57 | 7.5 | 17.9 | 3.2 |
| | Pc(%) | 0 | 87 | 41.5 | 31.7 | 0 | 76 | 36.8 | 28.9 | 1.1 |
| | Pa (%) | 0 | 4 | 1.2 | 1.6 | 0 | 8 | 1.7 | 2.7 | 0.7 |
| | Aae (%) | 0 | 5 | 3.3 | 2.6 | 0 | 50 | 12.7 | 14.4 | 0.3 |
| Hydrometric signatures | $I_{max}$ (mm/h) | 15 | 27 | 21.5 | 7.9 | 16 | 37.4 | 26.3 | 6.4 | 0.8 |
| | D (min) | 24 | 37 | 30.7 | 10.5 | 20 | 80.4 | 42 | 16 | 0.7 |
| | R (mm) | 808 | 2454 | 1508 | 726 | 411 | 6968 | 2237 | 1584 | 0.7 |
| | tp (min) | 38 | 127 | 84.0 | 37.8 | 54.0 | 217.5 | 102.9 | 54.0 | 0.8 |
| | tb (min) | 143 | 558 | 307.8 | 161.4 | 122.0 | 523.7 | 271.5 | 135.1 | 1.1 |
| | Φ(mm/h) | 15 | 27 | 21.3 | 7.7 | 16.1 | 35.4 | 25.4 | 5.6 | 0.8 |
| | Cr | 0.0 | 0.2 | 0.1 | 0.1 | 0.1 | 0.5 | 0.2 | 0.1 | 0.7 |
| | $Q_{mean}$ ($m^3/s$) | 0.5 | 2.7 | 1.3 | 0.9 | 0.1 | 1.6 | 0.7 | 0.5 | 1.8 |
| | $QS_{max}$ ($m^3/s:km^2$) | 0.3 | 0.9 | 0.5 | 0.3 | 0.1 | 2.7 | 0.9 | 0.8 | 0.6 |

**Table 5.** Statistics of cluster's attributes and signatures derived from Correlation distance. μ: mean; σ: standard deviation.

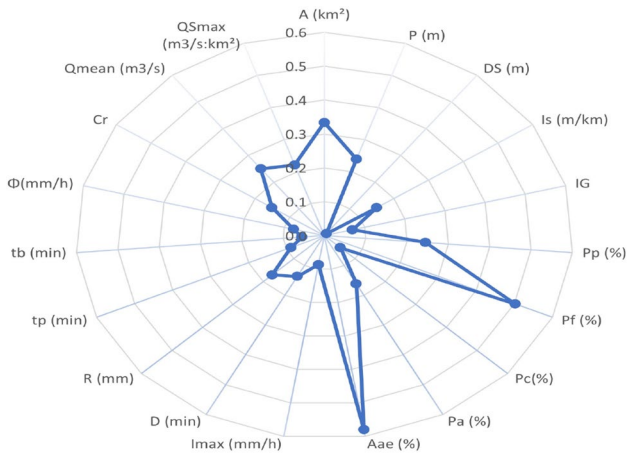



**Figure 4.** Variation of log absolute ratios of means attributes ($\rho = \mu_1/\mu_2$).

cover (Pf) and catchment's areas (A). Cluster1 is made of catchments with important forest cover, weak anti-erosive practices, and greater than 5 $km^2$. Cluster 2 is constituted by catchments with weak forest cover, important anti-erosive practices and smaller than 5$km^2$. Each cluster has its own specific hydrological behavior. This point will have to be respected in catchment modeling and runoff forecasting.

These results are in harmony with reviews detailing specific aspects of the hydrology of Mediterranean catchments such Mediterranean forest impact on catchment responses[42], the dryland hydrology[43] and erosion processes[44,45].

Finally, the delineation approach applied in current work reveals that distance between geomorphologic attributes and hydrometric signatures impacts the HCA delineation results so the hydrological pooled regions. This study can be considered an example case for Sud Mediterranean basins that can be extrapolated with other neighboring data as Algerian catchments.

## Conclusion

The current research described in this paper explores the use of unsupervised HCA in clustering Tunisian catchments which is applied with various distances calculated from associated attributes and signatures. Nineteen catchments are involved, and nine metric distances are explored to identify the most hydrologically similar clusters. Nineteen geomorphologic attributes and hydrometrical signatures (rainfall and flow signatures) are applied in this work to calculate diverse metric distances in HCA considered for delineating homogeneous clusters.

After performing the clustering step, Silhouette indexes are calculated for each cluster. They reveal that Correlation distance gives widely the most homogeneous clusters, compared with the other distances. It gives two clusters, not equally scattered (32% and 68% of total catchments) with average Silhouette indexes equal to 0.42 and 0.18.

Statistics show that the percentage of area affected by anti-erosive practices, the percentage of forest cover and catchment's area are the most discriminant attributes. However, hydrometrical signatures appear to be not relevant. This partitioning allowed to highlight two different hydrological behaviors which must be considered in modeling and/or forecasting.

Finally, these results can be helpful in regionalization strategy to calibrate hydrological models in south Mediterranean regions when the shortage of hydrometrical data is an occurring problem. They can be considered promising by the way that they can be advantageous in some cases of hydrologic predictions without need of heavy hydrologic information in ungauged catchments. Our study can be considered as a sample of Sud Mediterranean basins that can be extrapolated with data of other neighboring regions such as Algerian catchments.

## Data availability

All data generated or analyzed during this study are included in the published article " Gargouri-Ellouze E. & Bargaoui Z. Investigation with Kendall plots of infiltration index–maximum rainfall intensity relationship for regionalization. *Physics and Chemistry of the Earth,* Parts A/B/C, **34**(10–12), 642–653. (2009) " and its supplementary information files.

Received: 13 February 2023; Accepted: 11 July 2023
Published online: 26 July 2023

## References

1. Sivapalan, M. *et al.* IAHS decade on predictions in ungauged basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences. *Hydrol. Sci. J.* **48**, 857–880 (2003).
2. McGlynn, B. *et al.* A data acquisition framework for runoff prediction in ungauged basins. In *Runoff Prediction in Ungauged Basins: Synthesis across, Processes Places and Scales* (eds Blöschl, G. *et al.*) (Cambridge University Press, 2013).
3. Hrachowitz, M. *et al.* A decade of Predictions in Ungauged Basins (PUB)-a review. *Hydrol. Sci. J.* **58**(6), 1198–1255 (2013).
4. Viglione, A. *et al.* Comparative assessment of predictions in ungauged basins—Part 3: Runoff signatures in Austria. *Hydrol. Earth Syst. Sci.* **17**, 22632279 (2013).
5. Oudin, L., Kay, A., Andréassian, V. & Perrin, C. Are seemingly physically similar catchments truly hydrologically similar?. *Water Resour. Res.* **46**, W11558 (2010).
6. Singh, R., Archfield, S. A. & Wagener, T. Identifying dominant controls on hydrologic parameter transfer from gauged to ungauged catchments—A comparative hydrology approach. *J. Hydrol.* **517**, 985–996 (2014).
7. Srinivas, V. V., Tripathi, S., Ramachandra Rao, A. & Govindaraju, R. S. Regional flood frequency analysis by combining self-organizing feature map and fuzzy clustering. *J. Hydrol.* **348**(1–2), 148–166 (2008).
8. Rao, A. R. & Srinivas, V. V. *Regionalization of Watersheds—An Approach Based on Cluster Analysis Water Science and Technology Library* (Springer, 2010).
9. Burn, D. H. & Goel, N. K. The formation of groups for regional flood frequency analysis. *Hydrol. Sci. J.* **45**(1), 97–112 (2000).
10. Tsakiris, G., Nalbantis, I. & Cavadias, G. Regionalization of low flows based on canonical correlation analysis. *Adv. Water Resour.* **34**, 865–872 (2011).
11. Jared, D. *et al.* A watershed classification approach that looks beyond hydrology: application to a semi-arid, agricultural region in Canada. *Hydrol. Earth Syst. Sci.* **23**, 3945–3967 (2019).
12. Jain, A. & Dubes, R. *Algorithms for clustering data* (Prentice-Hall, Upper Saddle River, 1988).
13. Li, F. F. *et al.* Decomposition-ANN methods for long-term discharge prediction based on Fisher's ordered clustering with MESA. *Water Resour. Manage* **33**, 3095–3110 (2019).
14. Merz, R. & Blöschl, G. Regionalization of catchment model parameters. *J. Hydrol.* **287**(1), 95–123 (2004).
15. Goyal, M. K. & Gupta, V. Identification of homogeneous rainfall regimes, in northeast region of India using fuzzy cluster analysis. *Water Resour. Manag.* **28**, 4491–4511 (2014).
16. MacQueen J.B. Some methods for classification and analysis of multivariate observations. In *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, 1, 281–297 (University of California Press, Berkeley, 1967)
17. Peerzada, H. A. & Shilpa, D. Performance evaluation of clustering algorithm using different datasets. *J. Inf. Eng. Appl.* **5**(1), 39–47 (2015).

18. Durocher, M., Chebana, F. & Ouarda, T. Delineation of homogenous regions using hydrological variables predicted by projection pursuit regression. *Hydrol. Earth Syst. Sci.* **20**, 4717–4729 (2016).
19. Mirkin, B. *Clustering for Data Mining: A Data Recovery Approach* 1st edn. (Chapman and Hall/CRC, 2005).
20. Nathan, R. J. & Mc Mahon, T. A. Identification of homogeneous regions for the purposes of regionalization. *J. Hydrol.* **121**, 217–238 (1990).
21. Cunderlik, J. M., & Burn, D. H. Switching the pooling similarity distances: Mahalanobis for Euclidean. *Water Resour. Res*. **42–3** (2006).
22. Shirkhorshidi, A. S., Aghabozorgi, S. & Wah, T. Y. A comparison study on similarity and dissimilarity measures in clustering continuous data. *PLoSONE* **10**, 12 (2015).
23. Singla, S. *et al.* Evolution des relations pluie-débit sur des bassins versants du Maroc. In *World FRIEND Conference. Global Change: Facing risks and threats to water resources: Proceedings of the sixth world FRIEND conference* Vol. 25–29 (eds Servat, E. *et al.*) 679–687 (AISH, Wallingford, 2010).
24. Totz, S. *et al.* Winter precipitation forecast in the European and Mediterranean regions using cluster analysis. *Geophys. Res. Lett.* **44**(24), 12–418 (2017).
25. Ahattab, J., Serhir, N. & Lakhal, E. K. Vers l'élaboration d'un système d'aide à la décision pour la sélection des méthodes d'évaluation des pics de crue au Maroc: Réadaptation des méthodes classiques aux données hydrologiques récentes. *La Houille Blanche* **1**, 63–70 (2015).
26. Bargaoui, Z., Fortin, V., Bobée, B. & Duckstein, L. Une approche floue pour la détermination de la région d'influence d'une station hydrométrique. *Revue des Sciences de l'eau Eau* **11**(2), 255–328 (1998).
27. Gargouri-Ellouze E., & Bargaoui Z. Applicability of gamma-type geomorphologic instantaneous unit hydrograph on a small semi-arid sub catchment EGS-AGU-EUG joint assembly, 4288 (2003).
28. Bargaoui, Z. & Chebchoub, A. Investigations du caractère multifractal des débits maximaux annuels de crue/Investigation of multifractality of maximum annual flood discharges. *Hydrol. Sci. J.* **49**, 549–562 (2004).
29. Chérif, R. & Bargaoui, Z. Regionalization of maximum annual runoff using hierarchical and trellis methods with topographic information. *Water Resour. Manag.* **27**, 2947–2963 (2013).
30. Chérif R., & Gargouri-Ellouze E. Statistical regional index flood curve construction based on watersheds hydro-geomorphologic descriptors clustering. STAHY 10–11. Abu-Dhabi (2014)
31. Kotti, F. C. *et al.* Etude des pluies et des débits sur le bassin versant de la Medjerda, Tunisie Study of rainfall and discharges in the Medjerda watershed, Tunisia. *Bulletin de l'Institut Scientifique Rabat* **38**, 19–28 (2016).
32. Gargouri-Ellouze E., Chérif R., & Eslamian S. Geostatistics and flooding, homogeneous regions delineation for multivariate regional frequency analysis. In *Flood Handbook Analysis and Modeling*, Chapter CRC (Press Taylor and Francis Group, 2022).
33. Rousseeuw, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65 (1987).
34. Xiaogang, G. *et al.* An OD flow clustering method based on vector constraints: A case study for Beijing Taxi origin-destination data ISPRS. *Int. J. Geo-Inf.* **9**, 128 (1992).
35. Wazneh, H., Chebana, F. & Ouarda, T. B. M. J. Delineation of homogeneous regions for regional frequency analysis using statistical depth function. *J. Hydrol.* **521**, 232–244 (2014).
36. Gargouri-Ellouze, E. & Bargaoui, Z. Investigation with Kendall plots of infiltration index–maximum rainfall intensity relationship for regionalization. *Phys. Chem. Earth A/B/C* **34**(10–12), 642–653 (2009).
37. McIntyre N. *et al.* Ensemble predictions of runoff in ungauged catchments. *Water Resour. Res*. **41** (2005).
38. Sawicz, K. *et al.* Classification des bassins versants: analyse empirique de la similitude hydrologique basée sur la fonction des bassins versants dans l'est des États-Unis. *Hydrol. Earth Syst. Sci.* **15**, 2895–2911 (2011).
39. UNESCO. International glossary of hydrology. Preprint at https://library.wmo.int/doc_num.php?explnum_id=8209 (2012).
40. Ouarda, T. B. M. J., St-Hilaire, A. & Bobée, B. A review of recent developments in regional frequency analysis of hydrological extremes. *J. Water Sci.* **21**, 219–232 (2008).
41. Eden, J. *et al.* A global empirical system for probabilistic seasonal climate prediction. *Geosci. Model Dev.* **8**(12), 3947–3973 (2015).
42. Barnston, A. G. & Smith, T. M. Specification and prediction of global surface temperature and precipitation from global SST using CCA. *J. Clim.* **9**(11), 2660–2697 (1996).
43. Cosandey, C. *et al.* The hydrological impact of the Mediterranean forest: A review of French research. *J. Hydrol.* **301**(1–4), 235–249 (2005).
44. Cudennec, C., Leduc, C. & Koutsoyiannis, D. Dryland hydrology in Mediterranean regions—A review. Hydrological sciences. *J. Sci. Hydrol.* **52**(6), 1077–1087 (2007).
45. Shakesby, R. A. Post-wildfire soil erosion in the Mediterranean: Review and future research directions. *Earth Sci. Rev.* **105**(3–4), 71–100 (2011).
46. García-Ruiz, J. M., Nadal-Romero, E., Lana-Renault, N. & Beguería, S. Erosion in Mediterranean landscapes: changes and future challenges. *Geomorphology* **198**, 20–36 (2013).


## Author contributions

R.C. and G.E. wrote the main manuscript. figures2,3,4 are prepared by G.E., the others and tables are prepared by R.C. two authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-38608-6.

**Correspondence** and requests for materials should be addressed to R.C.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.