Catchments Classification: Multivariate Statistical Analysis for Physiographic Similarity in the Niger Basin

Jamilatou Chaibou Begou[1,2,*], Seifeddine Jomaa[3], Sihem Benabdallah[4], Pibgnina Bazie[2], Abel Afouda[1] and Michael Rode[3]

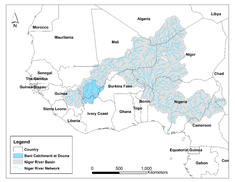
[1] Graduate Research Program (GRP) Climate Change and Water Resources, West African Science Service Centre on Climate Change and Adapted Land Use (WASCAL), Université d'Abomey-Calavi, Cotonou, Benin, jamilabegou@yahoo.fr, aafouda@yahoo.fr
[2] Centre Regional AGRHYMET, Niamey, Niger, p.bazie@agrhymet.ne
[3] Department of Aquatic Ecosystem Analysis and Management, Helmholtz Centre for Environmental Research – UFZ, Magdeburg, Germany, seifeddine.jomaa@ufz.de, michael.rode@ufz.de
[4] Centre de Recherche et des Technologies des Eaux (CERTE), Soliman, Tunisia, sihem.benabdallah@certe.rnrt.tn
[*] Correspondence: jamilabegou@yahoo.fr, Tel.: +227-9136-7854.

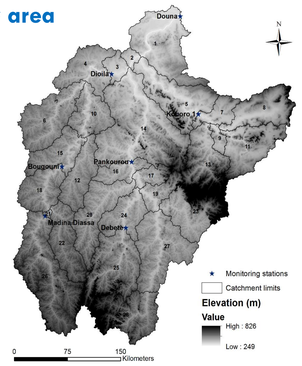
1. Introduction and Objectives

- Hydrological similarity between catchments is an essential concept in regionalization (Blöschl, 2001; Harman and Sivapalan, 2009; Wagener et al., 2007) and could be derived by a classification scheme.
- For a regionalization perspective, catchment classification consists in the search of hydrologically similar gauged catchment(s), from which hydrological information can be transferred to the ungauged catchment.
- **Objectives:**
- Determine a physiographic and climatic similarity between catchments located on the Bani basin.
- Determine the dominant factors that control this similarity.

2. Material and methods

a. Study area





- Catchment: Bani (Niger basin)
- Outlet: Douna
- Area: 100,000 Km^2
- Annual precipitation: 1050 mm
- Annual discharge: 184 m^3 s^{-1}

Figure 1. Location and Digital Elevation Model of the Bani basin.

b. Input catchments and catchments' attributes

- 28 nested catchments ranging in size from 92 km^2 to 10,910 km^2,
- 16 physiographic and climatic descriptors.

Table 1. Summary of catchment attributes (CAs).

Attribute	Description	Units
Slo1	Subbasin slope	%
Len1	Longest path within the subbasin	m
Sll	Field slope length	m
Csl	Subbasin tributary reach slope	m
Wid1	Subbasin tributary reach width	m
Dep1	Subbasin tributary reach depth	m
Lat	Latitude of the subbasin centroid	dd
Long	Longitude of the subbasin centroid	dd
Elev	Mean elevation of the subbasin	m
ElevMin	Minimum elevation of the subbasin	m
ElevMax	Maximum elevation of the subbasin	m
Shape_Leng	Subbasin perimeter	m
Shape_Area	Subbasin area	m^2
*P	Average annual precipitation on the subbasin	mm
AGRL	Proportion of Agricultural Land on the subbasin	%
Lf	Proportion of forest land on the subbasin	%

c. Hierarchical Clustering on Principal Components (HCPC)

- Methodology: based on the Hierarchical Clustering on Principal Components (HCPC) function proposed by (Husson et al., 2010);
- Performed under R package FactoMineR 1.28 (Husson et al., 2009; Lê et al., 2008);
- Combination of 2 exploratory data analysis methods: Principal Component Analysis (PCA) and Cluster Analysis (CA);
- Catchment attributes were first standardized;
- Appropriate number of Principal Components (PCs) chosen based on the Scree plot technique;
- Clustering built solely on the previously determined PCs;
- Distance between data points: Euclidean distance;
- Agglomerative method for merging two clusters: used the Ward's criterion

$$\sum_{k=1}^{K}\sum_{q=1}^{Q}\sum_{i=1}^{I_q}\left(x_{iqk}-\overline{x_k}\right)^2=\sum_{k=1}^{K}\sum_{q=1}^{Q}I_q\left(\overline{x_{qk}}-\overline{x_k}\right)^2+\sum_{k=1}^{K}\sum_{q=1}^{Q}\sum_{i=1}^{I_q}\left(x_{iqk}-\overline{x_{qk}}\right)^2,$$

Where x_{iqk} is the value of the variable k for the individual i of the cluster q, $\overline{x_{qk}}$ is the mean of the variable k for cluster q, $\overline{x_k}$ is the overall mean of variable k and I_q is the number of individuals in cluster q;

- Cluster description with the v-test

$$v-test=\frac{\overline{x}_{qk}-\overline{x}_k}{\sqrt{\frac{S^2}{n}\times\frac{N-n}{N-1}}}$$

Where n is the number of individuals in cluster q, N is the total number of individuals and S^2 is the variance of the dataset.

3. Results

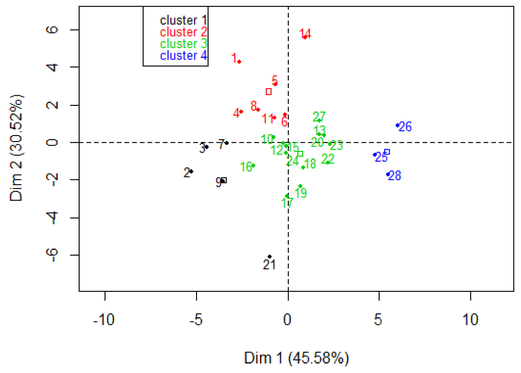


Figure 2. Hierarchical clustering representation on the map induced by the first 2 Principal Components on the Bani catchment.

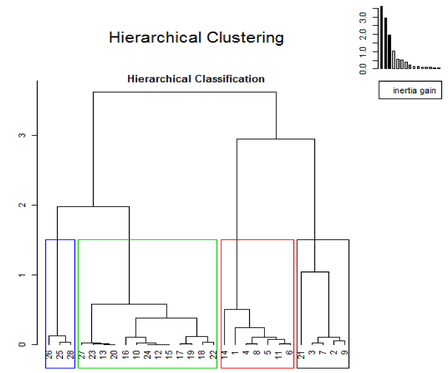


Figure 3. Hierarchical tree. Each rectangle represents a cluster of similar catchments.

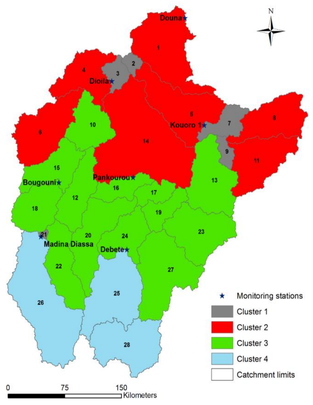


Figure 4. The spatial distribution of clusters of Physically-similar catchments on the Bani basin.

Table 2. Description of hierarchical clusters by the means of the v-test.

Variable	v-test	Mean in category	Overall mean	p-value
		Cluster 2		
AGRL	2.97	87.75	50.81	0.002936
Lat	2.76	12.17	11.41	0.005818
Lf	2.74	91.25	56.71	0.006213
ElevMin	- 2.68	265.57	285.96	0.007257
P	- 3.00	829.56	1011.18	0.002674
		Cluster 4		
Slo1	3.95	3.53	2.12	0.000080
ElevMax	3.41	801.67	559.00	0.000654
Elev	3.25	405.79	352.47	0.001171
ElevMin	2.77	321.00	285.96	0.005648
P	2.67	1279.97	1011.18	0.007662
Lat	- 3.16	9.94	11.41	0.001557

4. Conclusions

- The Bani basin was classified into 4 clusters of similar catchments (Figures 2-4),
- The topographic variables (Elev, ElevMin, ElevMax, Slo1), precipitation and the geographical position of the sub-catchment (Lat) were demonstrated to be the most important causes of similarity between catchments belonging to Cluster 2 and Cluster 4 (Table 2),
- This study permitted to propose the two nomenclature: Group of northerly flat and semi-arid catchments, and group of southerly hilly and humid catchments.

References

- Blöschl, G., 2001. Scaling in hydrology. Hydrological Processes, 15(4): 709-711.
- Harman, C., Sivapalan, M., 2009. A similarity framework to assess controls on shallow subsurface flow dynamics in hillslopes. Water Resources Research, 45(1): W01417.
- Husson, F., Josse, J., Lê, S., Mazet, J., 2009. FactoMineR: Multivariate Exploratory Data Analysis and Data Mining with R. R package version 1.12.
- Husson, F., Josse, J., Pages, J., 2010. Principal component methods - hierarchical clustering - partitional clustering: why would we need to choose for visualizing data?, Applied Mathematics Department, France.
- Lê, S., Josse, J., Husson, F., 2008. FactoMineR: An R Package for Multivariate Analysis. Journal of Statistical Software, 25(1): 1-18.
- Wagener, T., Sivapalan, M., Troch, P., Woods, R., 2007. Catchment Classification and Hydrologic Similarity. Geography Compass, 1(4): 901-931.

Acknowledgements: We thank **WASCAL** and **AGRHYMET**/**FFEM**/**CC** for providing funds to support this work which is part of a PhD Thesis.