RESEARCH ARTICLE OPEN ACCESS

Catchment classification: multivariate statistical analyses for physiographic similarity in the Upper Niger Basin

JamilatouChaibou Begou[1,2], Pibgnina Bazie[2] and Abel Afouda[1]
[1]Graduate Research Program (GRP) Climate Change and Water Resources, West African Science Service Centre on Climate Change and Adapted Land Use (WASCAL), University of Abomey-Calavi, Cotonou, Benin.
[2]Centre Regional AGRHYMET, PO Box 11011, Niamey, Niger.

Abstract
The objective of this study was to determine physiographic similarity, as indicator of hydrologic similarity between catchments located in the Upper Niger Basin, and to derive the dominant factors controlling each group singularity. We utilized a dataset of 9 catchments described by 16 physical and climatic properties distributed across a wide region with strong environmental gradients. Catchments attributes were first standardized before they underwent an integrated exploratory data analysis composed by Principal Component Analysis (PCA) followed by Hierarchical Clustering. Results showed a clear distribution into 2 major clusters: a group of easterly flat catchments and another of westerly hilly catchments. This nomenclature came from the interpretation of the main factors, topography and longitude, that seem to control the most important variability between both clusters. In addition, the hilly catchments were designated to be dominated by forest and ACRISOL soil type, two additional drivers of similarity. The outcome of this study can help understanding catchment functioning and provide a support for regionalization of hydrological information.
Keywords: catchments, Hierarchical Clustering, physiographic similarity, Principal Component Analysis, Regionalization.

I. Introduction

A core issue in hydrology is to make prediction of hydrological variable where it is not measured. This situation is of particular importance especially in developing countries where many river basins are ungauged [1-5]. This lack of information constraints water resources management and constitutes a stumbling block to adaptation to climate change in the sector of water hence increasing the vulnerability of rural population, particularly.

With the aim of predicting hydrological variables in ungauged basins, regionalization procedures are usually used. Different types of regionalization exist, and can be classified as [6] in: 1) regression methods, and 2) methods based on distance measures between gauged and ungauged sites. The former methods consist in deriving statistical relationships between catchment attributes and the optimized model parameters. Notwithstanding being considered as the most common regionalization approach for flow prediction in ungauged catchment [7], statistical methods are limited in use due to the presence of equifinality in calibrated model parameters. In fact, it becomes difficult to associate individual parameters with the physical characteristics of the catchment (each parameter can take several values) and thus, instead, complete parameter sets should be transferred to ungauged sites [8]. Another drawback of these methods is that most statistical models consider linearity between catchment attributes and model parameters [9, 10]. Consequently, in order to address the issue of model parameters non-uniqueness and propagate prediction uncertainty from gauged to ungauged catchment, similarity methods should be suitable.

Hydrologic similarity is an essential concept in regionalization [11-13]. Many similarity concepts have been proposed in the literature that attempt to represent various hydrologic processes occurring at different locations. [14], for instance, proposed three similarity concepts: spatial proximity, similar catchment attributes and similarity indices. In the first concept, catchments that are close to each other are assumed to behave hydrologically similarly. Geostatistical methods are based on this similarity measure. Many authors have indicated, for instance, the predominance of kriging methods on deterministic models in regions where the gauging network is sufficiently dense (e.g. [15, 16]). Nonetheless, it was pointed out that spatial proximity does not always involve functional similarity between catchments [17, 18], and thus[19, 20] suggest, instead, the application of hydrologically more meaningful distance measures. In the second concept, catchment attributes, such as catchment size, mean annual rainfall, and soil characteristics are used as indicators of physiographic similarity. Many studies stressed the value of parameter regionalization methods based on physiographic similarity, as a proxy for functional similarity ([10, 21, 22]. The third similarity concept is based on hydrologic function defined by similarity indices such as the aridity index of Budyko (e.g. [23, 24]), which has proved to be a valuable measure of catchment behavior.

Similarity of hydrological function between catchments could be derived by a classification scheme. As discussed by [13], the ultimate goal of classification is to understand the interaction between catchment structure, climate and catchment function. Additionally, [25]proposed four objectives of catchment classification which are: 1) nomenclature of catchments, 2) regionalization of information, 3) development of new theory, and 4) hydrologic implications of climate, land use and land cover change. Many authors attempted to classify catchments around the world into similar groups. For instance, [26]used 8 physiographic and meteorological variables to organize 21 catchments located within the Nile basin, into 2 homogeneous regions by applying a multivariate statistical analysis. In a different approach, [27]used self-organizing maps to classify around 300 Italian catchments according to several descriptors of the streamflow regime and geomorphoclimatic characteristics. As for [28], they distinguished only six dominant classes for 331 catchments across the continental United States using four similarity metrics. It is worth noting the work by[29]involving 24 worldwide large drainage basins, among which, the Niger basin. In fact, [29]considered sixteen geomorphologic and climatic variables into multivariate statistical analyses and obtained 6 clusters along with the description of the major controlling factors driving the hydrosedimentary response of each group. However, large river basins, as it is the case in [29], usually encompass several climatic regions and exhibit strong environmental gradients. Consequently, a global classification at such spatial scale can still hide significant internal heterogeneities among subcatchments, hence limiting our understanding of the hydrological functioning occurring at smaller catchments. Therefore, it is essential to break down the scale and provide more detailed classification scheme, and this is essential especially when prediction in small ungauged catchments is foreseen. However, only one a priori classification of the Niger basin exists and have been proposed by the Niger Basin Authority (e.g. [30]) which subdivided the whole basin into 4 physio-climatic regions: the Upper Niger, the Niger Inner Delta, the Middle Niger, andthe Lower Niger. Nevertheless, this classification falls short of providing a quantitative assessment of the degree of (dis)similarity within and between the so-called homogenous regions.

In the light of these examples, the main objective of this study was to classify subcatchments of the Upper Niger into similar groups according to their physio-climatic parameters. The specific objectives were to: 1) reduce the dimension of the input dataset containing catchment attributes by a Principal Component Analysis, and 2) perform a hierarchical clustering of subcatchments based on the reduced dataset. This study provides the first ever quantification of similarity among catchments with respect to physiographic characteristics on a large tropical river basin at finer spatial scale. Nor descriptors, neither statistics themselves are actually novel in the broad literature, but their combined use in that particular area to evaluate the gain of homogeneity with increasing number of clusters, is.The questions that will be addressed in this study were: (i) can the Upper Niger further be separated into similar groups of catchments based on their physical characteristics, and if so, (ii) what are the dominant controls of similarity between catchments.

II. Material and methods

2.1. The study area

The present study was conducted within the Upper Niger (Fig. 1). This basin is composed by to mutually independent subbasins: the Upper Niger subbasin controlled by the Koulikoro gauging station and the Banisubbasin at the Douna outlet, each covering an area of 120,000 km2 and 101,000 km2, respectively. The study area is shared by four West African countries: Guinea, Cote d'Ivoire, Mali and Burkina Faso, in a lesser extent. To avoid confusion, the parent basin is called the Upper Niger and its subbasin upstream Koulikoro is called the Upper Niger subbasin.

Altitudes are unevenly distributed across the Upper Niger. The extreme west and south of the basin are hilly zones. The Tinkissosubcatchment, for instance, is situated in the Fouta-Djalon Mountain, which culminates at more than 1000 m in the region of Dabola[31]. Similarly, the south of the basin is shaped into plateau and mountains, the most important of which is situated between Milo and Dion rivers and reaches its highest point at 1500 m. In contrast, the Bani watershed's topography is gently sloping, with altitudes ranging from 249 m to 826 m. Average annual precipitation (period 1981-2000) varies from 1500 mm y-1 in the humid Guinean zone in the south-west (region of Kissidougou) to 620 mm y-1 in the Sahelian zone in the North-east (region of Segou). The vegetation is dominated by the presence of closed evergreen forest in the highlands of the Fouta Djalon Mountain, whereas the Baniis mainly the domain of savannah with small spots of deciduous forest. ACRISOL is the most important soil grouping on the majority of subcatchments, except at the subcatchments controlled by Bougouni and Kouro1 gauging stations.

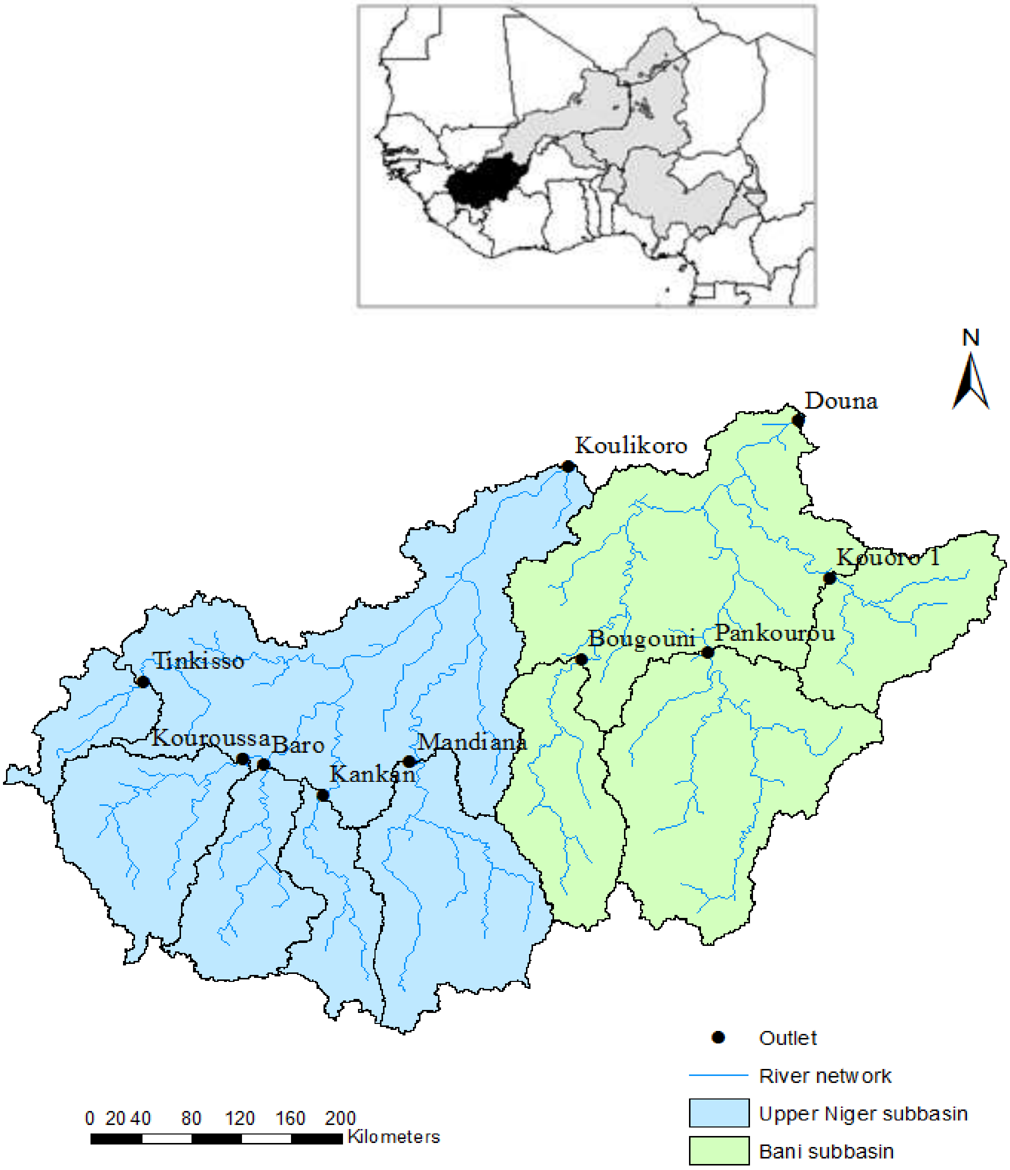

Figure 1: localization of the Upper Niger basin and the study catchments

2.2 Catchments and catchments' attributes

A total of 9 candidate catchments were selected and range in size from 6379 km^2 to 101,456 km^2 and were hereinafter given the name of their corresponding outlet. For example, Bougouni referred to the subcatchment controlled by the Bougouni outlet. Three of them are included in the Bani, while five are located on the Upper Niger subbasin, (Fig. 1), and are referred to as Group I and Group II, respectively. The Dounasubbasin, which is actually the Bani, is the biggest catchment and was added on purpose to test similarity across spatial scale. In addition to belonging to two hydrologically non-connected subbasins, Group I and Group II individuals were chosen to be non-nested sites in order to provide a better structure of independence between subcatchments. Furthermore, these subcatchments have not been affected by anthropogenic activities able to significantly modify their flow regime and have been chosen to be located in the headwaters of both subbasins.

This study make the implicit assumption that the physical similarity based on the selected catchment attributes, is a proxy of hydrological functioning of a catchment. Therefore the choice of catchment attributes (CAs) is of great importance. Selected CAs are related to the shape (e.g. area, length) and the topography (e.g. slope, elevation) of each subbasin and its main tributary reach and were derived by application of the SWAT model (at watershed delineator and HRU analysis processing steps required for SWAT model setup). The same input spatial data (Table 1) were used to characterize Group I and Group II subcatchments. The selection of the appropriate CAs can also depend on the physical meaning of the model parameters (Mps) that will subsequently be involved in information regionalization. For instance, in the SWAT model, the curve number parameter (CN2) which is considered among the most sensitive Mps, depends on the soil and land use characteristics of the catchment [32]. Therefore, two other characteristics related to land use and soil were considered as descriptors: Forest and ACRISOL. Forest represents the proportion of area covered by forest, and ACRISOL gives information about the soil texture based on the relative proportion of sand, silt and clay. As ACRISOL remains the dominant soil in the majority of the study catchments, its proportion is used to indicate the presence of more than 35% of clay in each catchment. Forest and ACRISOL were calculated using the following equations:

$$\text{Forest} = \left(\frac{Af}{A}\right) \times 100 \quad (1)$$

$$\text{ACRISOL} = \left(\frac{Aacs}{A}\right), \quad (2)$$

Where*Af* is the area covered by forest within a watershed, *Aacs* is the area covered by ACRISOL, and A is the total area of the watershed.

Last, it is very common to use climatic characteristics such as long-term annual precipitation as indicator of similarity. Thus, average annual precipitation was computed for each subcatchment on the period 1981-2000. A detailed description of the 16 CAs is given in Table 2.

Table 1:Input data for SWAT model to derive catchments attributes on the Upper Niger basin.

Data type	Description	Resolution/period	Source	Processing
Topography	Conditioned DEM	90 m	USGS hydrosheds[a]	SWAT Watershed Delineator
River	River network	500 m	USGS Hydrosheds[a]	SWAT Watershed Delineator
Land use/cover	GLCC version 2	1 km	Waterbase[b]	SWAT HRU Analysis
Soil	FAO Soil Map	Scale 1:5000000	FAO [c]	SWAT HRU Analysis
Precipitation data	Rainfall	Daily/1981-2000	AGRHYMET	Arithmetic mean

[a]http://hydrosheds.cr.usgs.gov
[b]http://www.waterbase.org
[d]http://www.fao.org/geonetwork

Table 2:Summary of catchment attributes derived by the SWAT model as input for multivariate statistical analyses on the Upper Niger basin.

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	-
Long	Longitude of the subbasin centroid	-
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
[a]P	Average annual precipitation on the subbasin (mm)	mm
Forest	Proportion of forest on the subbasin	%
ACRISOL	Proportion of ACRISOL on the subbasin (%)	%

[a] Calculated on the period 1981-2000

2.3 Multivariate statistical analyses

Multivariate statistics used in this study are Principal Components Analysis (PCA) and Cluster Analysis (CA), and were performed under R package FactoMineR[33, 34], version 1.28.

PCA and CA are frequently used in hydrological studies [25, 26,35], and commonly applied in a pre-processing of a set of variables prior to the classification, to provide a convenient lower-dimensional summary of thedataset, or as a classification tool itself. PCA reduces a dataset containing a large number of variables to a dataset containing fewer new variables that are linear combinations of the original ones. These linear combinations are chosen to represent the maximum possible fraction of the variability contained in the original data and are called Principal Components (PCs). CA attempts to separate observations into groups of similar characteristics called clusters.

The methodology utilized in this study was based on the Hierarchical Clustering on Principal Components (HCPC) function proposed by [36]. This method combines three exploratory data analysis methods, Principal Component methods, Hierarchical Clustering and partitioning, to improve data analysis.The chosen Principal Components method is the PCA, because retained CAs are quantitative variables. PCA was used herein as a pre-process for clustering, i.e., the hierarchical clustering is solely built on the determined PCs. In that case, the clustering is more stable than the one obtained from original variables [36]. Input variables, i.e., CAs, were standardized because they are not measured on comparable scales. The appropriate number of PCs was chosen based on the scree plot technique [37]. Then, a hierarchical agglomerative clustering was performed on the PCs previously determined. The measure of distance between data points was based on the Euclidean distance (the same was used in PCA) and the agglomerative method for merging two clusters used the Ward's criterion. According to this criterion, the total inertia (variability) is decomposed in within-group and between-group inertia, and the pair of groups to be merged is chosen that minimizes the growth of within-group inertia. Equation (3) gives the formula for calculating the total inertia of a dataset:

$$\sum_{k=1}^{K}\sum_{q=1}^{Q}\sum_{i=1}^{I_q}\left(x_{iqk}-\overline{x_k}\right)^2=\sum_{k=1}^{K}\sum_{q=1}^{Q}I_q\left(\overline{x}_{qk}-\overline{x_k}\right)^2+\sum_{k=1}^{K}\sum_{q=1}^{Q}\sum_{i=1}^{I_q}\left(x_{iqk}-\overline{x_{qk}}\right)^2, \quad (3)$$

Total inertia = Between-group inertia + Within-group inertia

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.

The last step consists in choosing the appropriate number of clusters when it is not preassigned, that is, the stopping point of clustering that maximizes similarity within clusters and maximizes dissimilarity between clusters. HCPC function suggests an "optimal" number Q of clusters when the decrease in within-group inertia between Q - 1 and Q is from far greater than the one between Q and Q + 1 (see [36]for a thorough description of the HCPC function). Results of HCPC function can be presented in different ways: (1) a factor map, which displays results of the hierarchical clustering on the map induced by the first PCs, (2) a 2-dimensional dendrogram or hierarchical tree, and (3) a 3-dimensional dendrogram in which the hierarchical tree is incorporated into the factor map. The latter representation can solely be used to get an integrated visualization of the dataset. However, dispersion of data points is somehow masked in that way. Therefore, the factor map was presented in the results section for a better visualization of individuals' dispersion on the plan formed by PCs, while the hierarchical treeoffers a good insight of the variability increase between clusters.

III. Results

3.1 Catchments clustering

It is interesting to briefly describe the intermediate result of PCA. It permitted to determine 2 PCs that explain 81.33% (63.84% for Dim1 and 17.49% for Dim2) of the total variance of the original data set. The subsequent clustering was then performed on these PCs. Results are presented on Fig. 2 and Fig. 3.

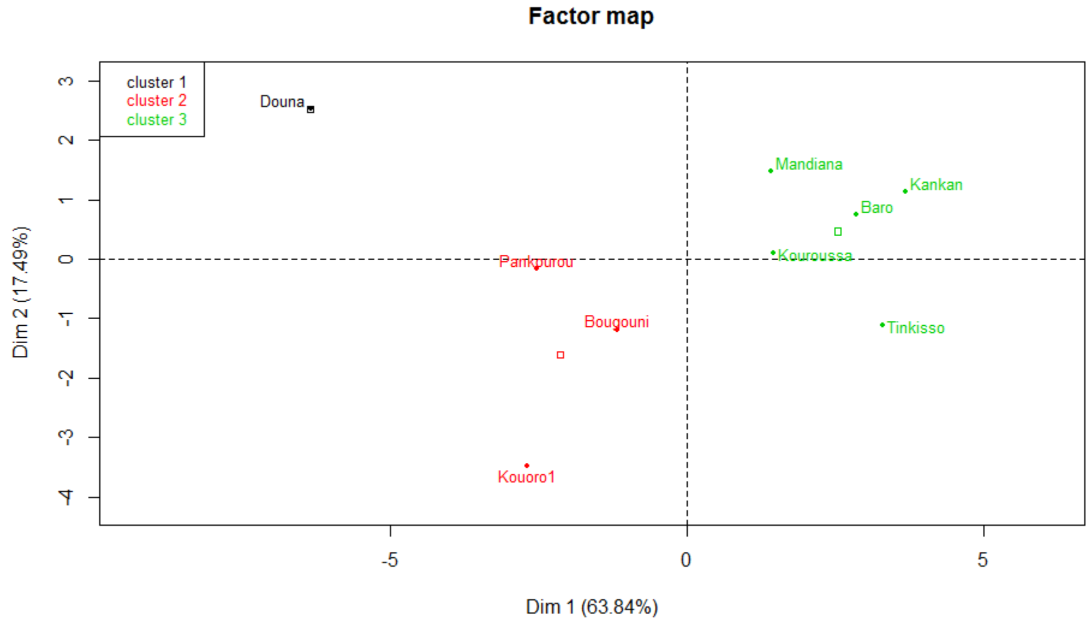

Figure 2: Hierarchical clustering representation on the map induced by the first 2 Principal Components on the Upper Niger basin. Catchments are colored according to the cluster they belong to, the barycenter of each cluster is represented by a square and Dim1 and Dim2 are the first two Principal Components on which the hierarchical clustering is built.

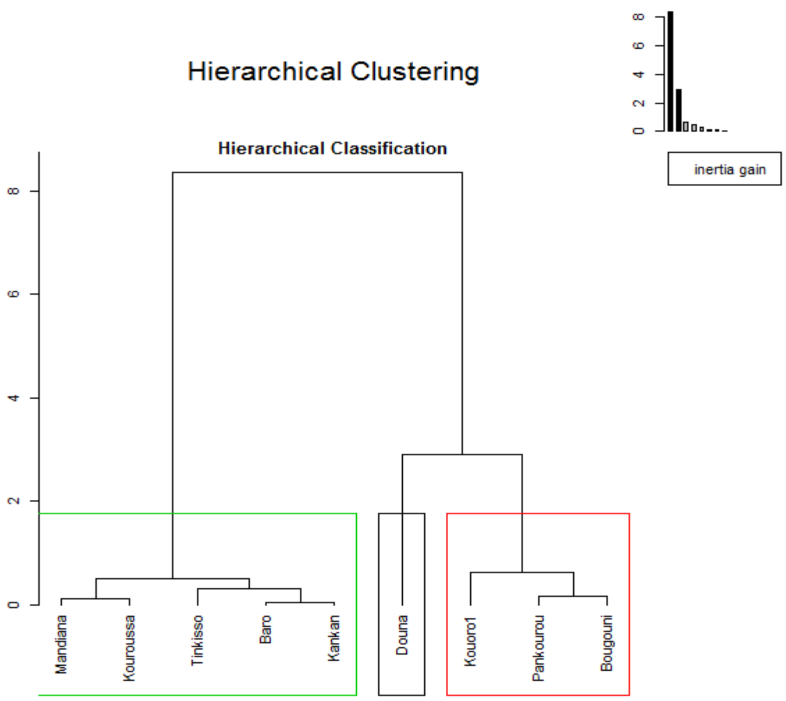

Figure 3: hierarchical clustering of the Upper Niger catchments. On the hierarchical classification or tree, each rectangle represents a cluster of similar catchments. The barplot(inertia gain) gives the decrease of within-group variability with increasing number of clusters.

Table 3: Description of hierachical clusters. In bold, positive v.test value indicating that the variable has a value greater than the overall mean, and in italic, negative v.test value indicating that the variable has a value smaller than the overall mean . All v.test values are significant at the probability p = 0.05

Variable	v.test	Mean in the category	Overall mean	p-value
Cluster 2				
Long	**2.01**	- 6.43	- 8.32	0.045
Elev	*- 1.97*	376.51	454.59	0.048
Cluster 3				
Elev	**2.63**	520.42	454.59	0.0084
ElevMin	**2.58**	344.6	314	0.0097
Slo1	**2.48**	5.6	4.21	0.0131
Forest	**2.35**	79.01	52.54	0.0186
ElenMax	**2.34**	1219.2	1029.78	0.0194
Csl	**2.24**	0.18	0.13	0.0252
Acrisol	**2.01**	62.06	46.36	0.0442
Long	*- 2.49*	- 9.81	-8.32	0.0126

IV. Discussion and conclusions

Overall, results of this study showed that the Upper Niger can be classified into 2 major clusters of similar catchments based on physiographic characteristics. In addition, topographic variability and geographical position of the subcatchment were demonstrated to exert a stronger control on separating clusters, and permitted to propose a kind of nomenclature of clusters: the group of easterly flat catchments assigned to the Bani, and the one of westerly hilly catchments, assigned to the Upper Niger subbasin. The latter is further characterized by the dominance of Forest and ACRISOL as the major soil type. These results expectedly answer the questions posed at the beginning of this work. However, due to limited availability of literature on this area, it is difficult to show how these results fit in with existing knowledge on that topic. A broader comparison can only be made about the dominant controls on similarity in different contexts. For instance, [26]demonstrated that topographic parameters (e.g., mean stream slope, minimum elevation, and maximum elevation) provide the major categorization of catchments of the equatorial Nile, and proposed the same nomenclature of flat and hilly regions. Likewise, [29]showed that the whole Niger basin is close to the group of basins characterized by topographic parameters (hypsometry and mean elevation), which can be considered as the major driving forces of its hydrosedimentary response.

Nonetheless, it is important to note that no cluster analysis can produce a definitive classification because the results are depending on the dataset used and other kind of subjective choices (choice of classification algorithm and distance metric, [25]). It is also acknowledged that the actual limitation that arose within this study was the absence of geologicaldescriptors, limiting thus our understanding of subsurface controls. In spite of the limitations discussed above, these are encouraging results, showing on one hand the relevance of physical characteristics to give information about the spatial dissimilarity characterizing a large tropical river basin, and on the other, the value of statistical analyses (such as the HCPC function) as a pertinent tool for exploring similarity among catchments. Concerning the assumption of correspondence between physical and functional similarity made in this study, [18]pointed out that this assumption may not always be verified. Further studies can try to find out its validity in the present case study, by evaluating, for instance, the performance of a regionalization method to transfer information within and between clusters. The use of other similarity concepts (such as similarity indices) applied to the same catchments could also give a good platform of discussion.

V. Acknowledgements

This study was funded by the German Ministry of Education and Research (BMBF) through the West African Science Service Centre on Climate Change and Adapted Land Use (WASCAL; www.wascal.org) that supports the Graduate Research Program Climate Change and Water Resources at the University of Abomey-Calavi. We would like also to thank AGRHYMET Regional Centre for providing additional funds through the French Global Environment Facility (FFEM/CC) project.

References

[1.] H. Bormann, B. Diekkrüger, Possibilities and limitations of regional hydrological models applied within an environmental change study in Benin (West Africa), Physics and Chemistry of the Earth, Parts A/B/C, 28(33–36), 2003, 1323-32.

[2.] E. Kapangaziwiri, D.A. Hughes, T. Wagener, Incorporating uncertainty in hydrological predictions for gauged and ungauged basins in southern Africa, Hydrological Sciences Journal, 57(5), 2012, 1000-19.

[3.] D. Mazvimavi, A.M.J. Meijerink, H.H.G. Savenije, A. Stein, Prediction of flow characteristics using multiple regression and neural networks: A case study in Zimbabwe, Physics and Chemistry of the Earth, Parts A/B/C, 30(11–16), 2005, 639-47.

[4.] M. Minihane. Estimating mean monthly streamflow in the lugenda river, northern mozambique. Putting prediction in ungauged basins into practice: Canadian Water Resources Association; 2013. p. 185-96.

[5.] P. Ndomba, F. Mtalo, A. Killingtveit, SWAT model application in a data scarce tropical complex catchment in Tanzania, Physics and Chemistry of the Earth, Parts A/B/C, 33(8–13), 2008, 626-32.

[6.] M. Hrachowitz, H.H.G. Savenije, G. Blöschl, J.J. McDonnell, M. Sivapalan, J.W. Pomeroy, B. Arheimer, T. Blume, M.P. Clark, U. Ehret, F. Fenicia, J.E. Freer, A. Gelfan, H.V. Gupta, D.A. Hughes, R.W. Hut, A. Montanari, S. Pande, D. Tetzlaff, P.A. Troch, S. Uhlenbrook, T. Wagener, H.C. Winsemius, R.A. Woods, E. Zehe, and C. Cudennec, C., A decade of Predictions in Ungauged Basins (PUB)—a review. Hydrological Sciences Journal, 58 (6), 2013, 1–58.

[7.] T. Wagener, H.S. Wheater, Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty, Journal of Hydrology, 320(1–2), 2006, 132-54.

[8.] A. B¨¢rdossy, Calibration of hydrological model parameters for ungauged catchments, Hydrology and Earth System Sciences Discussions, 2006.

[9.] R. Merz, G. Blöschl, Regionalisation of catchment model parameters, Journal of Hydrology, 287(1–4), 2004, 95-123.

[10.] J. Parajka, R. Merz, G. Blöschl, A comparison of regionalisation methods for catchment model parameters, Hydrol Earth SystSci, 9(3), 2005, 157-71.

[11.] G. Blöschl, Scaling in hydrology, Hydrological Processes, 15(4), 2001, 709-11.

[12.] C. Harman, M. Sivapalan, A similarity framework to assess controls on shallow subsurface flow dynamics in hillslopes, Water Resources Research, 45(1), 2009, n/a-n/a.

[13.] T. Wagener, M. Sivapalan, P. Troch, R. Woods, Catchment Classification and Hydrologic Similarity, Geography Compass, 1(4), 2007, 901-31.

[14.] G. Blöschl, Scaling and Regionalization in Hydrology, Elsevier; 2011.

[15.] G. Laaha, J.O. Skøien, G. Blöschl, Spatial prediction on river networks: comparison of top-kriging with regional regression, Hydrological Processes, 28(2), 2014, 315-24.

[16.] J.O. Skøien, G. Blöschl, Spatiotemporal topological kriging of runoff time series, Water Resources Research, 43(9), 2007, n/a-n/a.

[17.] G. Ali, D. Tetzlaff, C. Soulsby, J.J. McDonnell, R. Capell, A comparison of similarity indices for catchment classification using a cross-regional dataset, Advances in Water Resources, 40, 2012, 11-22.

[18.] L. Oudin, A. Kay, V. Andréassian, C. Perrin, Are seemingly physically similar catchments truly hydrologically similar?, Water Resources Research, 46(11), 2010, n/a-n/a.

[19.] A. Bárdossy, G.G.S. Pegram, L. Samaniego, Modeling data relationships with a local variance reducing technique: Applications in hydrology, Water Resources Research, 41(8), 2005, n/a-n/a.

[20.] Y. He, A. Bárdossy, E. Zehe, A review of regionalisation for continuous streamflow simulation, Hydrol Earth SystSci, 15(11), 2011, 3539-53.

[21.] P.F. Dornes, B.A. Tolson, B. Davison, A. Pietroniro, J.W. Pomeroy, P. Marsh, Regionalisation of land surface hydrological model parameters in subarctic and arctic environments, Physics and Chemistry of the Earth, Parts A/B/C, 33(17–18), 2008, 1081-9.

[22.] I. Masih, S. Uhlenbrook, S. Maskey, M.D. Ahmad, Regionalization of a conceptual rainfall–runoff model based on similarity of the flow duration curve: A case study from the semi-arid Karkheh basin, Iran, Journal of Hydrology, 391(1–2), 2010, 188-201.

[23.] M. Sivapalan, M.A. Yaeger, C.J. Harman, X. Xu, P.A. Troch, Functional model of water balance variability at the catchment scale: 1. Evidence of hydrologic similarity and space-time symmetry, Water Resources Research, 47(2), 2011.

[24.] S. Tekleab, S. Uhlenbrook, Y. Mohamed, H.H.G. Savenije, M. Temesgen, J. Wenninger, Water balance modeling of Upper Blue Nile catchments using a top-down approach, Hydrol Earth SystSci, 15(7), 2011, 2179-93.

[25.] K. Sawicz, T. Wagener, M. Sivapalan, P.A. Troch, G. Carrillo, Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA, Hydrol Earth SystSci, 15(9), 2011, 2895-911.

[26.] J.M. KileshyeOnema, A.E. Taigbenu, J. Ndiritu, Classification and flow prediction in a data-scarce watershed of the equatorial Nile region, Hydrol Earth SystSci, 16(5), 2012, 1435-43.

[27.] M. Di Prinzio, A. Castellarin, E. Toth, Data-driven catchment classification: application to the pub problem, Hydrol Earth SystSci, 15(6), 2011, 1921-35.

[28.] E. Coopersmith, M.A. Yaeger, S. Ye, L. Cheng, M. Sivapalan, Exploring the physical controls of regional patterns of flow duration curves – Part 3: A catchment classification system based on regime curve indicators, Hydrol Earth SystSci, 16(11), 2012, 4467-82.

[29.] J. Raux, Y. Copard, B. Laignel, M. Fournier, N. Masseï, Classification of worldwide drainage basins through the multivariate analysis of variables controlling their hydrosedimentary response, Global and Planetary Change, 76(3–4), 2011, 117-27.

[30.] ABN, Atlas Bassin du Niger (ABN: Niamey, 2007).

[31.] Y. Brunet-Moret, P. Haoeron, J.P. Lamagat, M. Molinier, Monographiehydrologique du fleuveNiger.Tome 1-Niger superieur (ORSTOM: Paris, 1986).

[32.] H. Sellami, Uncertaintyassessment of hydrological impacts of climatic change in smallmediterraneancatchments, doctoral diss.,UniversitéCatholique de Louvain, Belgique, 2014.

[33.] F. Husson, J. Josse, S. Lê, J. Mazet. FactoMineR: Multivariate Exploratory Data Analysis and Data Mining with R. R package version 1.12, 2009. Available from: http://factominer.free.fr.

[34.] S. Lê, J. Josse, F. Husson, FactoMineR: An R Package for Multivariate Analysis, Journal of Statistical Software, 25(1), 2008, 1-18.

[35.] A. Daigle, A. St-Hilaire, D. Beveridge, D. Caissie, L. Benyahya, Multivariate analysis of the low-flow regimes in eastern Canadian rivers, HydrolSci J, 56(1), 2011, 51-67.

[36.] F. Husson, J. Josse, J. Pages, Principal component methods - hierarchical clustering - partitional clustering: why would we need to choose for visualizing data?,technical report (Agrocampus: France, 2010).

[37.] I.T. Jolliffe, Principal Component Analysis (Springer, 2002).