



AGU Books

**Towards a framework of catchment classification for hydrologic predictions and water resources management in the ungauged basin of the Congo River: An a priori approach**

Journal:	AGU Books
Manuscript ID	2020-Feb-CH-1205.R1
Wiley - Manuscript type:	Chapter
Date Submitted by the Author:	n/a
Complete List of Authors:	<p>Tshimanga, Raphael; na, Congo Basin Water Resources Research Center -CRREBaC, University of Kinshasa          Bola, Gode; University of Kinshasa, Department of Natural Resources Management; Congo Basin Water Resources Research Center          Kabuya, Pierre; Rhodes University, Institute for Water Research; Congo Basin Water Resources Research Center , natural Resources Management          Nkaba, Landry; University of Kinshasa, Congo Basin Water Resources Center &amp; Dept. Natural Resources Management          Neal, Jefferey; University of Bristol          Hawker, Laurence; University of Bristol          Trigg, Mark; University of Leeds, Civil Engineering          Bates, Paul; University of Bristol, School of Geographical Sciences          Hughes , Denis; Rhodes University,          Laraque, Alain; GET-UMR CNRS/IRD/UPS – UMR 5562 du CNRS, UR 234 de l'IRD – 14, av. E. Belin – 31400 Toulouse, France.          Woods, Ross; University of Bristol Faculty of Engineering, Civil Engineering          Wagener, Thorsten; University of Bristol Faculty of Engineering, Civil Engineering</p>
Primary Index Term:	1804 - Catchment < 1800 - HYDROLOGY
Index Term 1:	1719 - Hydrology < 1700 - HISTORY OF GEOPHYSICS
Index Term 2:	1615 - Biogeochemical cycles, processes, and modeling (0412 , 0414 , 0793 , 4805 , 4912) < 1600 - GLOBAL CHANGE
Index Term 3:	1874 - Ungaged basins < 1800 - HYDROLOGY
Index Term 4:	1873 - Uncertainty assessment (1990 , 3275) < 1800 - HYDROLOGY
Keywords:	Catchment classification, Complexity, Hydrologic prediction, Socio-hydrology, Ungauged basin, Water resources management
Abstract:	The Congo Basin exhibits tremendous heterogeneities, out of which it emerges as an intricate system where complexity will vary consistently over time and space. Increased complexity in the absence of adequate knowledge will always result in increased uncertainties. One way of

	<p>simplifying this complexity is through an understanding of organisational relationships of the landscape features, which is termed here as catchment classification. The need for a catchment classification framework for the Congo Basin is obvious given the basin's inherent heterogeneities, the ungauged nature of the basin, and the pressing needs for water resources management that include the quantification of current and future supplies and demands, which also encompass the impacts of future changes associated with climate and land use, as well as water resources operational policies. The need is also prompted by many local-scale management concerns within the basin. This study uses an a priori approach to determine homogenous climatic-physiographic regions that are expected to underline dominant hydrological processes characteristics. A set of 1740 catchment units are partitioned across the whole basin, based on a set of comprehensive criteria, including natural break of the elevation gradient (199 units), inclusion of socio-economic and anthropogenic systems (204 units), and water management units based on traditional nomenclature of the rivers within the basin (1337 units). The identified catchment units are used to assess existing datasets of the basin physical properties, necessary to derive descriptors of the catchments characteristics. An unsupervised classification, based on Hierarchical Agglomerative Cluster algorithm is used, that yields 11 homogenous groups that are consistent with the current perceptual understanding of the Congo Basin physiographic and climatic settings. These regions represent therefore an a priori classification that will be further used to derive functional relationships of the catchments, necessary to enable hydrological prediction and water management in the basin.</p>

## **Towards a framework of catchment classification for hydrologic predictions and water resources management in the ungauged basin of the Congo River: An a priori approach**

Raphael M. Tshimanga<sup>1</sup>, Gode Bola<sup>1</sup>, Pierre Kabuya<sup>1,2</sup>, Landry Nkaba<sup>1</sup>, Jeffery Neal<sup>3</sup>, Laurence Hawker<sup>3</sup>, Mark A. Trigg<sup>4</sup>, Paul Bates<sup>3</sup>, Denis A. Hughes<sup>2</sup>, Alain Laraque<sup>5</sup>, Ross Woods<sup>6</sup>, Thorsten Wagener<sup>6</sup>

1. Congo Basin Water Resources Research Center (CRREBaC) & Dpt. Natural Resources Management, University of Kinshasa, Box 117 KIN XI, DR Congo
2. Institute for Water Research, Rhodes University, Box 94, Grahamstown 6140, South Africa
3. School of Civil Engineering, University of Leeds, Leeds LS2 9JT, UK
4. School of Geographical Sciences, University of Bristol, UK
5. GET-UMR CNRS/IRD/UPS – UMR 5562 du CNRS, UR 234 de l'IRD – 14, av. E. Belin – 31400 Toulouse (France)
6. Department of Civil Engineering, University of Bristol, UK

[raphtm@yahoo.fr](mailto:raphtm@yahoo.fr)

### **ABSTRACT**

The Congo Basin exhibits tremendous heterogeneities, out of which it emerges as an intricate system where complexity will vary consistently over time and space. Increased complexity in the absence of adequate knowledge will always result in increased uncertainties. One way of simplifying this complexity is through an understanding of organisational relationships of the landscape features, which is termed here as catchment classification. The need for a catchment classification framework for the Congo Basin is obvious given the basin's inherent heterogeneities, the ungauged nature of the basin, and the pressing needs for water resources management that include the quantification of current and future supplies and demands, which also encompass the impacts of future changes associated with climate and land use, as well as water resources operational policies. The need is also prompted by many local-scale management concerns within the basin. This study uses an a priori approach to determine homogenous climatic-physiographic regions that are expected to underline dominant hydrological processes characteristics. A set of 1740 catchment units are partitioned across the whole basin, based on a set of comprehensive criteria, including natural break of the elevation gradient (199 units), inclusion of socio-economic and anthropogenic systems (204 units), and water management units based on traditional nomenclature of the rivers within the basin (1337 units). The identified catchment units are used to assess existing datasets of the basin physical properties, necessary to derive descriptors of the catchments characteristics. An unsupervised classification, based on Hierarchical Agglomerative Cluster algorithm is used, that yields 11 homogenous groups that are consistent with the current perceptual understanding of the Congo Basin physiographic and climatic settings. These regions represent therefore an a priori classification that will be further used to derive functional relationships of the catchments, necessary to enable hydrological prediction and water management in the basin.

**Key words:** Catchment classification, Complexity, Hydrologic prediction, Socio-hydrology, Ungauged basin, Water resources management.

## 1. Introduction

The landscape can be viewed as being organized in structural units that have a spatial pattern. The dynamic interactions of these structural units determine the functional characteristics of the landscape features. The flow patterns, transfer pathways, catchment response to events and inputs such as natural and human induced disturbances, and the spatial and temporal distribution of such inputs (Harvey & Gonzalez Villalobos, 2007), remain under the influence of the landscape elements in relation to each other (Bull *et al.*, 2003; Lexartza-Artza and Wainwright, 2009). As a result, this conditions the processes in a non-linear manner (van Oost *et al.*, 2000), as well as imposing some longitudinal, lateral, vertical and temporal behaviours on the catchment processes (Ward *et al.*, 2002). Description of structural and functional characteristics of the landscape units can then provide a basis for understanding of processes across scales and ease our efforts for hydrological predictions and management of water resources under conditions of environmental change (Beven *et al.*, 2019).

The Congo River Basin (CRB) is framed within 10°N, 12°E to 14°S, 34°E (corner-corner coordinates, Figure 1), and shares the drainage divides with other large river and lake basins of Africa, including the Zambezi basin in the south, Nile basin in the east and Lake Chad basin (through Chari River) in the north, all of which drain in opposite directions, thus emphasizing the complexity of regional drainage patterns. Within the CRB itself, there is no single classical mode of drainage pattern which emphasizes the heterogeneous nature of the underlying geology. The CRB is approximately circular, ringed by topographic highs, also called “Swells” (Kadima *et al.*, 2011) or “Rises” (Runge, 2008), with a large central depression, commonly known as the Cuvette Centrale. These topographic highs include the Atlantic Rise on the western ridge of the basin, the Asante Rise on the northern ridge, the Mitumba Mountains on the eastern ridge, and the Lunda Rise (Shaba and Angola swells) in the southern ridge. The main rivers generated from the Atlantic Rise are known as the Sangha, Likouala Mossaka, Kouyou, Alima, Lefini, Nkeni, Foulakari, Mponzo, Luozi, Inkisi. The Asante Rise encompasses the streams that drain the CRB starting from its drainage divides with the Chari (Lake Chad) and Nile Basins to the main trunk of the Congo River. This drainage unit is known as the Oubangui (or Ubangui) basin, the name of the main stream that connects all the upstream north-eastern tributaries to the main trunk of the Congo River. The Mitumba Mountains mark a clear drainage divide between the Congo and the Nile basins, and constitute the source of the main rivers that provide substantial flow contribution to the Congo River at Kisangani and the lower part of Lualaba. The southern rim of the CRB is flanked by the Lunda Rise, which is shaped by the Angolan highlands in the south-west and the Shaba, or north Zambian swell in the south-east. The main rivers of the southern CRB, notably the Kasai and Lualaba Rivers, rise in the highlands of the Angolan and the Shaba Swells (Luanda Rise), respectively. An internal drainage basin, the *Cuvette Centrale*, characterizes the central part of

the Congo Basin (Laraque *et al.*, 1998; Thieme *et al.*, 2005; Bwangoy *et al.*, 2010; Munzimi *et al.*, 2019). Traditionally, the nomenclature system of these rivers and their tributaries refer to tribes or customary practices of water use and management (e.g. rituals of initiation stages, water quality, river depth, hydrological events, etc.) which underlines the role of first-order classification based on social history, and has a meaning to local people. The rivers generated from the above main physiographic units cross areas of various heights, slopes, vegetations, soils and geologies before discharging their average annual flow volume of about 1300 billion m<sup>3</sup> into the Atlantic Ocean, thus covering an area of about 3.7 million km<sup>2</sup> (Figure 1). There are tremendous heterogeneities, out of which the CRB emerges as an intricate system where complexity will consistently vary over time and space. Increased complexity in the absence of adequate knowledge will always result in increased uncertainties. One way of simplifying this complexity is through an understanding of organisational relationships of the landscape features, which is termed in this study as catchment classification.

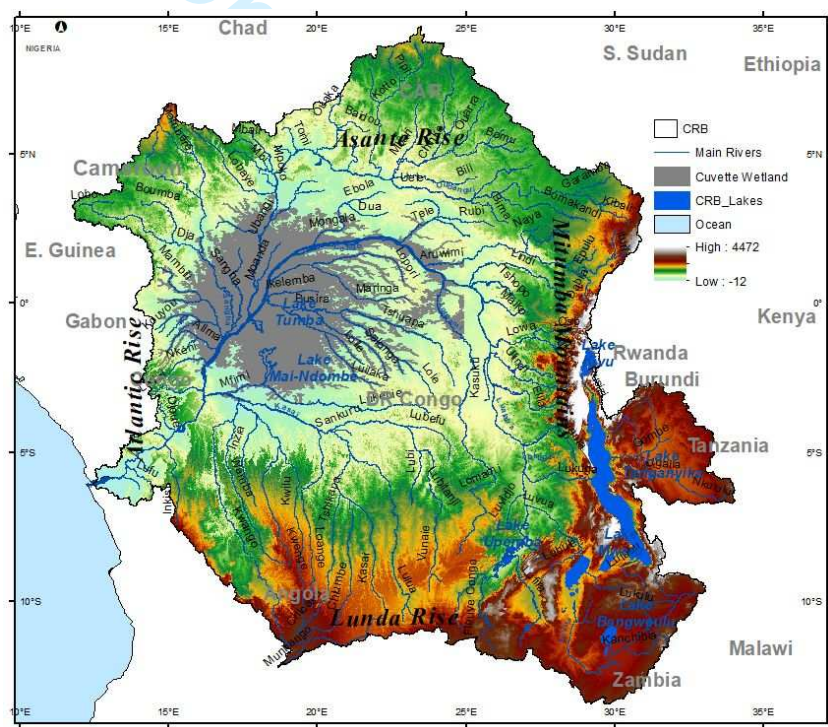


Figure 1 Physiographic setting of the Congo River Basin

Carrillo *et al.* (2011) state that catchment classification is an efficient way that helps synthesize our understanding of the interaction between climate variability, catchment characteristics (e.g. vegetation, soils, topography) and the resulting hydrological response. It is also a crucial step in improving predictions in ungauged basins. According to Sivakumar *et al.* (2015), the need for a catchment classification framework in hydrology was realized as early as in the 1930s, and since

then several attempts have been made to advance the concept. In recent years, the concept of catchment classification emerged with an emphasis on the need to understand organizational relationships of the landscape features; to uncover potential process theories that are embodied within natural heterogeneities of the landscape form, structure and functions (Sivapalan, 2005; Wagener *et al.*, 2007; Sivakumar *et al.*, 2015); to address the challenge of generalizing knowledge derived from local observations and understanding of the catchment response characteristics, as well as the underlying process controls; to provide guidance for measurement and modelling, and provision of constraints for predictions in ungauged basins and estimate of impacts of environmental changes.

The need for a catchment classification framework for the CRB is obvious given the tremendous heterogeneities outlined above, the ungauged nature of the basin, and the pressing needs for water resources management that include the quantification of current and future supplies and demands as well as the impacts of future changes associated with climate and land use. The need is also prompted by many local-scale management concerns within the basin. There is a need to sustain optimal use of the basin water resource services that encompass hydropower, water supply, fisheries, agriculture, transportation, and maintenance of aquatic ecosystems under the conditions of environmental changes. A previous attempt to provide a framework of catchment classification for the CRB (Tshimanga, 2012) identified functional relationships based on pre-defined areas of dominant slopes and elevations, but those relationships were fraught with issues due to spatial resolution of the assessment units, and were not strong enough to provide a convincing argument for the inter-dependency between the hydrological response groups and their physiographic controls. This study uses an a priori approach to determine homogenous climatic-physiographic regions that are expected to underline dominant hydrological processes characteristics. These regions represent therefore an a priori classification that will be further used to derive relationships of the catchment structures, processes and functions, necessary to enable hydrological prediction and water management in the basin, which also encompass implications for comprehensive studies such as eco-hydrology, water pollution, flood management, and operational water policies.

## **2. A conceptual framework of catchment classification for the Congo River Basin**

The basic tenet of catchment classification is as noted by Brown *et al.* (2014), the ability to provide a solid foundation towards assessing the impact of natural flow variability, associated ecological conditions and management of water resources across a range of scales. It provides a tool for developing inventories, data interpretation, interpolating information from specific sites to larger areas, strategic development of objectives and standards, design of monitoring strategies, and evaluating the uniqueness of areas. Wagener *et al.* (2008) stress the need for a catchment classification system that accounts for uncertainty, has a predictive power rather than

just being descriptive, and includes catchment services and disservices, which would explicitly link hydrology to current and future societal issues.

In this regard, it is worth defining key components that would explicitly guide a framework of catchment classification of a river basin given the level of its complexity and potentials. In this study, the key components that are considered necessary to develop a framework of catchment classification for the CRB include *partition of landscape and definition of catchment units; catchment structures; catchment processes; and functional relationships* of the catchment structures and processes, which determine the *various uses and services of water resources* and are also under the *influence of human induced disturbances* (Figure 2).

### **2.1 Landscape partition and definition of catchment units**

As previously mentioned, the most commonly known partition of the landscape into hydrologically meaningful units within the CRB is made of the major drainage units of the Oubangui (North-East), Sangha (North-West), Kasai (South-West) and Lualaba (South-East) Rivers, as well as the *Cuvette Centrale* that characterizes the central part of the basin. These draining units are very large and heterogenous, with their areas ranging from at least 21 2400 km<sup>2</sup> (Sangha River) to 961 648 km<sup>2</sup> (Lualaba River), and could hardly be used to address neither water issues nor the impacts of environmental change at the appropriate scales of management. Therefore, a comprehensive partition of the landscape units to define catchments for hydrological prediction and water management would benefit the purpose of catchment classification under the study. It should be noted that definition or partition of comprehensive catchment units is a prerequisite to any processes of catchment classification, none of which has been previously established for the CRB.



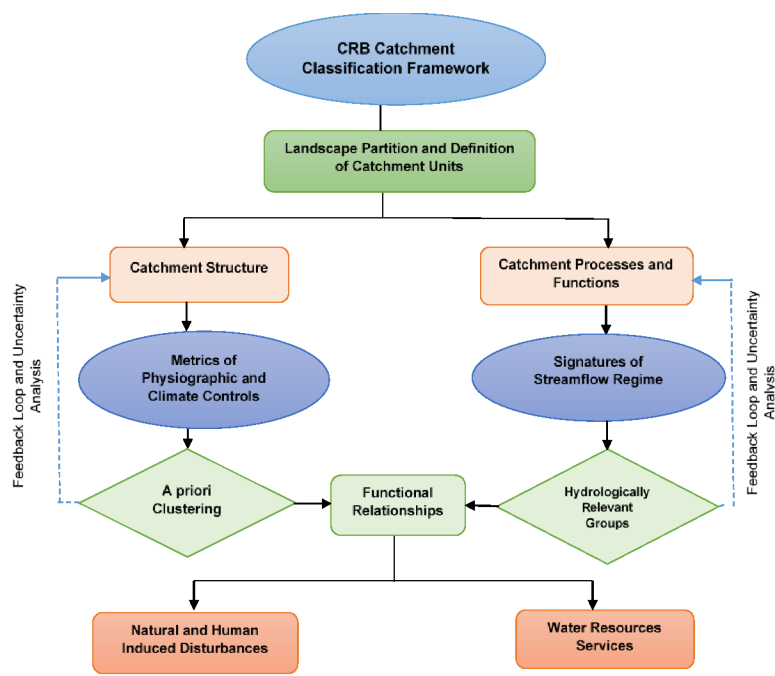


Figure 2 A conceptual framework of catchment classification for the CRB

### 2.2 Catchment structure

Consider the water balance at the land surface where  $\Delta S/\Delta t$  represents the change in storage over time. Two major concepts can be defined from this expression, where the storage is composed of (1) *media properties* with (2) *variation* in space and time. Media properties are structural characteristics where the material properties vary from point to point (Poehls & Smith, 2009). Variation in the structural characteristics determines the heterogeneity of the physiographic settings. Heterogeneity is often related to physical features of the natural system, such as topography, soil characteristics, geology and vegetation (Wigmosta & Prasad, 2005). The complexity of the hydrological processes is related to the heterogeneity of the landscape properties, which in turn are compounded by temporal and spatial variability at all scales (Sivapalan, 2005). The need for a holistic approach, rather than a fragmented description of the landscape heterogeneities, has also supported the increasing recognition of natural and multi-scale heterogeneities. This need has further propelled investigations for a so-called “new unified theory of hydrology at the catchment scale” (Sivapalan, 2005; Sivakumar & Singh, 2012). The previewed innovative unified theory attempts to address multi-scale heterogeneities as a natural and intrinsic part of the catchment hydrology, as well as discovering new catchment scale processes in relation to the patterns of heterogeneity. At the same time, the theory attempts to identify the interconnections and feedback between patterns and processes over a range of scales, and their interpretations in terms of their functions.



In the CRB, the complexity of hydrological processes is partly due to scale issues (Bloschl & Sivapalan, 1995; Woods, 2005) associated with the large geographic area covered by the basin ( $3.7 \times 10^6 \text{ km}^2$ ), and partly due to the sparse sources of information on physical basin properties, climate drivers, and observed hydrological response (see Tshimanga, 2012; Tshimanga & Hughes, 2014 for further reading about the hydrological complexity in the CRB). Understanding the organisational relationships of the landscape features would assist the development of a coherent framework of catchment classification and water resources management decisions in the CRB. The main assumption driving this first level *a priori* criteria is that the physical basin attributes such as climate, topography, vegetation, soil types, and geology exert a large control on the basin hydrological response, and thus areas with similar physiographic characteristics may lead to similar hydrological responses (e.g. Laraque *et al.*, 2009). Sivakumar *et al.* (2015) provide an overview of environmental metrics that have been globally used in studies of catchment classification. Therefore, it should be possible to use common principles of diagnostic evaluation to make inferences about physical basin properties and derive useful implications for the dominant runoff generation processes in the CRB.

### 2.3 Catchment processes and functional characteristics

If we consider a natural system as an entity, it is therefore obvious to understand that its behaviour as a whole depends upon the individual contributing units (elements), and their interactions. As previously mentioned, the dynamic interactions of the landscape structural units determine the functional characteristics of the landscape features such as the flow patterns, transfer pathways, catchment response to events and inputs such as natural and human induced disturbances, and the spatial and temporal distribution of such inputs. An initial development of a semi distributed rainfall-runoff model for the CRB (Tshimanga *et al.* 2011; 2012; Hughes *et al.* 2013; Tshimanga & Hughes, 2014), identified a number of hydrological processes inherent to catchment functioning in the context of the CRB. These processes include surface and sub-surface response to rainfall at the small scale, but also include storage and attenuation effects of wetlands, floodplains, natural lakes and the channel systems of large rivers. These processes occur differently across the basin and are under the influence of the various physical factors. Furthermore, the hydrodynamic processes of wetland-channel connectivity (Yuan *et al.*, 2017; Carr *et al.*, 2019) as well as source to sink processes of biogeochemical matters (Borges *et al.*, 2019) are considered very relevant as part of catchment functioning in the CRB.

## 2.4 Water resources services

Key decisions in river basin management concern options to meet the needs for development and economic growth while maintaining the environmental carrying capacity. Many factors beyond the immediate technical and economic considerations can have huge impacts on the functioning of a natural system and thus jeopardise services provided through available water resources (human services and ecosystem services). The CRB generates an annual flow volume of about 1 300 billion m<sup>3</sup> at its outlet, which represents about 40% of the African continent's discharge (Crowley *et al.*, 2006). The basin holds a huge potential for water resources development with multiple goods and services that include hydro-power, water supply, fisheries, agriculture, transportation, and maintenance of aquatic ecosystems. A few pre-feasibility studies highlighted potential sites for the development of more than 40 000 MW of continuous electrical power production (Maher, 1994; Mukheibir, 2007). Opportunities to achieve a further 100 000 MW are also underlined. The CRB is located over a tropical humid region with difficulty to ensure proper maintenance of road infrastructure. In these conditions, the river network provides some 25 000 km of inland navigation that connect the riparian countries internally but also to international markets. Since the colonial period, this river network has served the need for exchange of good and services through navigation. The river system, its wetland and floodplain along with the existence of tropical humid forests constitute one of the major eco-regions of the world that sustains some thousands of endemic species of the aquatic biodiversity (Thieme *et al.*, 2005). Availability of water for domestic uses, agriculture and mining is an essential asset for water security and economic growth in the region. Beside these potentials for multiple goods and services provided by catchment water resources, there is a range of problems related to water resources management, such as analysis of quantity and quality aspects of the catchment runoff processes, reservoir system operation, groundwater development and protection, climate-land use change, water resources allocation for various uses, and river restoration, among others (Partow, 2011). The question of how the many services offered by water resources of the CRB should be optimized (Wagener *et al.*, 2008) is at the heart of this catchment classification framework.

## 3. Methodological approach

### 3.1 Definition of landscape units

In this study, the partition of the landscape into meaningful hydrological and water management units or catchments involves a three-stage procedure, including the partition based on natural break of the elevation gradient, inclusion of anthropogenic features, and characterization of water management units. The reasoning behind the choice of this three-level criterion is also given below.

### 3.1.1 Natural break of topographic gradients

The land surface topography constitutes the natural criteria of catchment delineation, and topographic ridges are usually taken as their boundaries (Wagener *et al.*, 2008). Various studies have demonstrated the role of the topographic gradient (elevation-slope) in defining the magnitude-frequency relationships of geomorphologic processes (e.g. Hovius *et al.*, 2000). Garbrecht & Martz (1999) mention the importance of catchment slope for runoff, erosion and energy fluxes. Flow paths and travel distances within the catchment are subject to the variation and forms of slopes. Many characteristics of the catchment hydrological yield, such as the available kinetic energy for downstream outflow (Mazimavi, 2003), the runoff, and base flow responses (Vogel & Kroll, 1992), are related to the terrain slope. Elevation-area and slope-area relationships can explain many characteristics of the landforms, while their derivatives, such as hypsometric curve, hypsometric integral, wetness index, circularity, slope frequency, slope position and many others, can be used to explain a variety of lithologic and hydro-climatic conditions of the catchment hydrological functions. With the advent of DTMs that provide terrain information at pixel size, it has become possible to derive topographic attributes at finer scales. By examining the frequency distribution of elevations in the CRB (Figure 3), it is possible to derive homogenous classes of basin elevation, which would represent areas of the most frequent elevation. In this study, a frequency distribution of the elevation gradient (Figure 3) is used to determine dominant classes of the landform for the CRB, based on the recently released global MERIT-DEM (Yamasaki *et al.*, 2017).

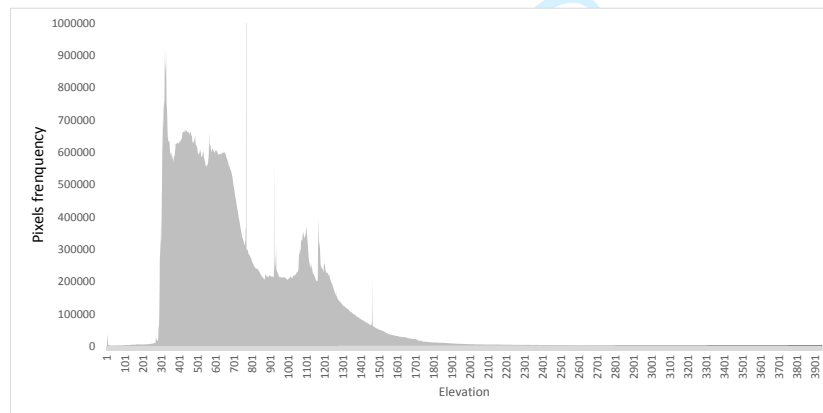


Figure 3 Frequency distribution of the elevation structure in the CRB from MERIT DEM

Jenks natural break optimization (De Smith *et al.*, 2018) was then applied to discriminate major landforms across the elevation gradient, which helped derive classes that represent the dominant areas that are frequent across the elevation gradient in the CRB. The Natural break classification has the ability to group classes based on break points that are relatively large jumps

in data values, and can be used to minimize variance within classes and maximize the variance between classes (Park & Kim, 2019). The approach also focused on the use of a threshold value that generates stream orders consistent with the observed river network from headwaters to the outlet located at the Atlantic Ocean. The search for a threshold value consistent with the primary river network of the basin was guided based on a topographic map of surface water of the CRB produced from field survey during the colonial period (IRB, 1950). This topographic map also refers to the traditional nomenclature of the rivers, which was used to define the water management units (see section 3.1.3). Currently, the D8 algorithm is the most appropriate and widely used method to extract drainage network and its related contributing units (Mayorga et al., 2005), and was used in this study under the ArcGIS toolbox. Through trial and error, a threshold accumulation value of 3000 (number of contributing areas needed to generate a true stream) was obtained that is consistent with a stream order of level nine (see Figure 4), based on Strahler classification (Strahler, 1954). Digitization and burning of river networks in the DEM were performed for places such as urban areas where rivers appeared to be distorted.

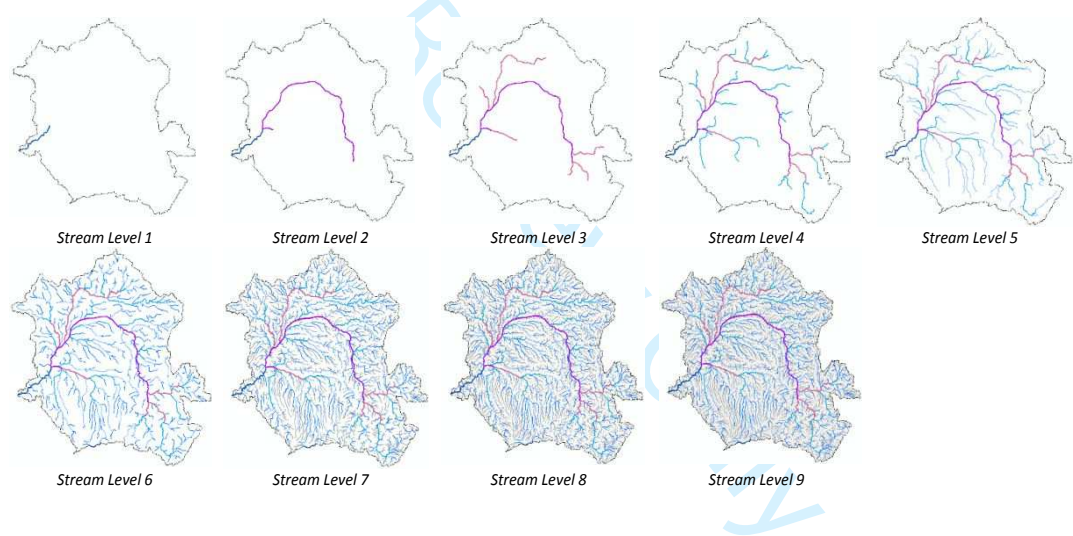


Figure 4      Hierarchical levels of rivers in the CRB based on Strahler order

Delineation points were generated for the various landforms that were differentiated from the frequency distribution of the elevation gradient in the basin, out of which the partition of catchment units was obtained. The classified features of the landforms are visually recognized and can be approximately delineated on the basis of a river network upon which the delineation points are positioned at places where sudden changes appear across the elevation gradient. These changes may include the exit/entering points from/to wetland, lake or reservoirs, headwater areas and different incremental zones along the river network. The Figure 5 shows an excerpt of the application of the methodology for the catchment of Pool Malebo. The identified

units are therefore considered representative of the main landforms of the CRB, including flat topography, valleys, ridges, etc.

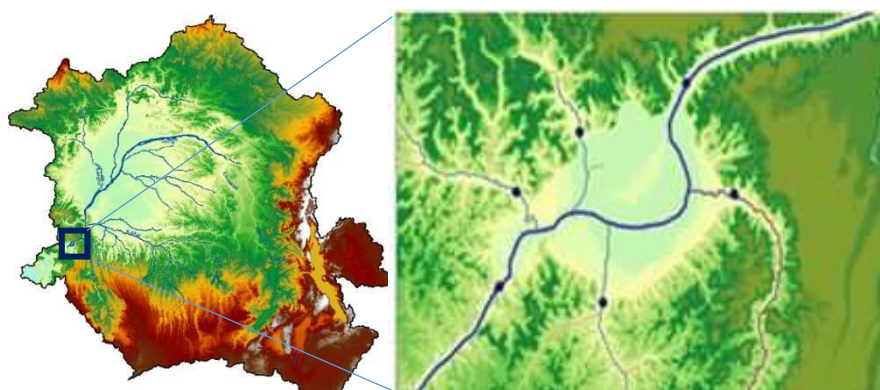


Figure 5 An excerpt of the application of the approach for the Pool Malebo where the delineation points are positioned at the outlets and inlets of the contributing area. Blue dots represent the delineation points.

The validation was carried out using information from the topographic map of the CRB (IRB, 1950) and the Google Earth Engine (Gorelick *et al.*, 2017). This topographic map (1: 5 000 000) is among the few materials that were produced during the colonial period with accurate details, and it provides information such as stream network from sources to the mouth at the Atlantic Ocean, the names of the streams, streamflow measuring sites, cities, cascades, hydropower sites, navigable waterways, wetlands, etc. This map was therefore digitised, geo-referenced, and used to validate the accuracy of river network processed from MERIT DEM data.

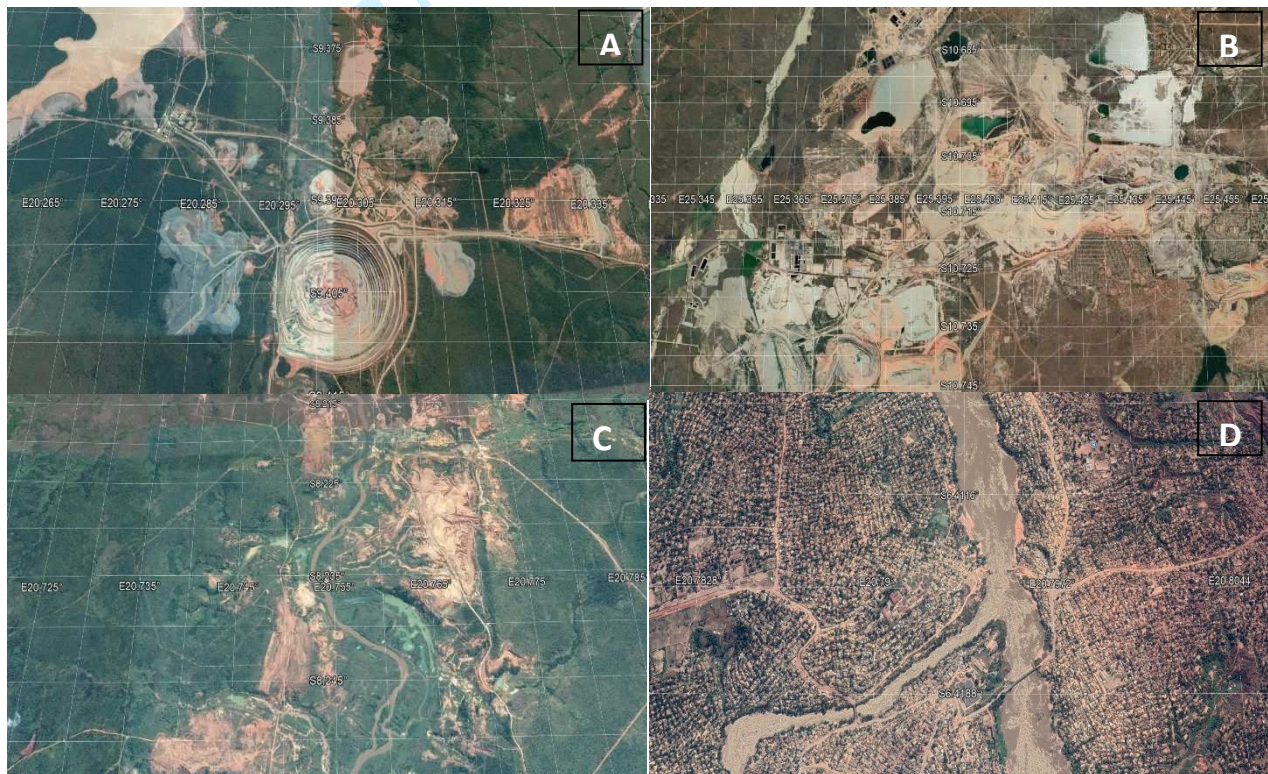
### 3.1.2 Inclusion of anthropogenic features

The past century has seen a strong demographic expansion, with critical anthropogenic implications on hydrologic systems, including through urbanization, land use and land cover change, water pollution and climate change. Therefore, the study of catchment responses to anthropogenic influences and their implications for catchment classification has become extremely important (Sivakumar *et al.*, 2015). Wagener *et al.* (2008) state the importance of including the implication of the values of different signatures for humans and ecosystems in addition to the signatures that describe hydrologically relevant characteristics of the catchment response. This inclusion is necessary to create a classification framework that explicitly defines the societal relevance of catchment behaviour and that is directly significant for water resources management.

The second level of catchments definition in this study involves inclusion of areas of urban settlements, mining activities as well as other socio-economic features of interest (Figure 6).



During the last two decades, there has been massive movement and concentration of population due to civil unrest in remote areas of the CRB, thus leading to expansion of cities and rapid-uncontrolled urbanization, which have huge impacts on land use patterns and thus the hydrologic regime of the basin. Besides, the potential of the basin in mineral resources is subject to large scale mining activities (gravel and alluvial), which are the main sources of soil erosion and river sedimentation. Mining activities in the CRB have resulted in dislocation of river networks due to informal diversions of river channels especially through artisanal exploitation of the mineral resources (Figure 6). In this study, we used a database of mining areas (CAMI, 2012) that we plotted in Google Earth to screen potential areas of significant mining activities. Figure 6 illustrates cases of anthropogenic features that were identified as relevant for inclusion in catchment units.



**Figure 6** Features of mining and urban areas included in catchments delineation (on top two industrial mining of Saurimo in Angola (A- catchment area: 7050 km<sup>2</sup>) and Kimoto in Katanga (B- catchment area: 5617 km<sup>2</sup>); bottom left an artisanal mining with dislocation of the river system in Camissombo, upper Kasai basin (C- catchment area: 14285 km<sup>2</sup>); and bottom right major town of Tshikapa in DRC (D- catchment area: 9189 km<sup>2</sup>)).

Notwithstanding the impacts of deforestation and rainfed agriculture in the basin, it appears therefore that these types of land use are not localized, and thus difficult to include as a criterion for catchment definition. In addition to the urban settlement and main mining areas, the port

infrastructure for navigation and main gauging sites were added to the definition of the catchments. The port areas are mainly located on the main channel of the basin where the navigation traffic is important. This aspect is included in this delineation framework to account for the hydrodynamic processes of the main channels for those reaches. These channel reaches delimited within the sub-basins will help monitoring of water level (navigation early warning signals) or flow routing along the port channels through hydrodynamic modelling. A consideration is also given to account for the evaluation of the hydrological simulations that will be generated from the resulting sub-basins. To this end, the geographical coordinates of the water monitoring infrastructure have been considered. Where they are too close or overlap the natural downstream identified point of delineation, they are prioritized and used to generate a catchment.

### 3.1.3 Characterization of water management units.

The two first levels of catchment partitions identified in this study are relevant for assessing the performance of predictive models, and land use and change across the major physiographic domains of the CRB. However, they will likely hardly address water management issues such as water pollution and flood control, local sediment transport processes, environmental water requirements, and water resources system operation policies, etc., that are of daily challenge at smaller scales in the CRB (Partow, 2011). An example case is illustrated in Figure 7 for the unit of Pool Malebo near the urban city of Kinshasa where the partition integrates many rivers that variably present challenges of management including urban water supply and sanitation, flooding, river encroachment, sediment, fishery, navigation, etc. One way of addressing this challenge is to identify water management units with regard to socio-economic and environmental interests. While the approaches to the partition of water management units could be legion, in this study, a criterion of partitioning based on the existing river names has been found relevant (Figure 7, right). Figure 8 provides an illustration of the river names that have been identified for a South-East area of the CRB. The justification is that many of the river's names in the CRB are in relation with customary practices and possess local meaning. In fact, rivers in the CRB bear the names that relate them to tribes, customary practices of water use and management such as rituals of initiation stages, water quality, river depth, social or major hydrological events, etc. In essence, every river that has a name also has a history of management, which brings in the concept of social hydrology recently advocated in the PUB decade (Sivapalan *et al.*, 2003). In this study, 1339 river names were identified for the whole CRB through assessment and digitisation of colonial archive maps as well as interview with local communities. These 1339 rivers were then used to delineate the corresponding catchments nested within the 403 major physiographic domains.



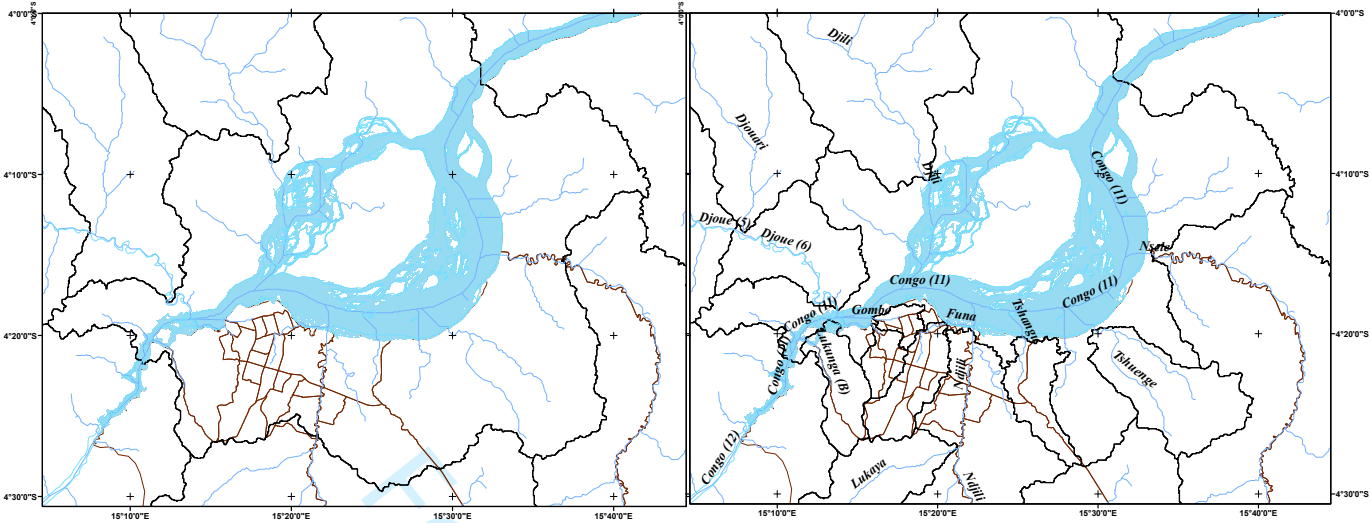


Figure 7 Illustration of second (Left) and third (Right) levels catchments delineation for the unit of Pool Malebo near the city of Kinshasa in DRC.

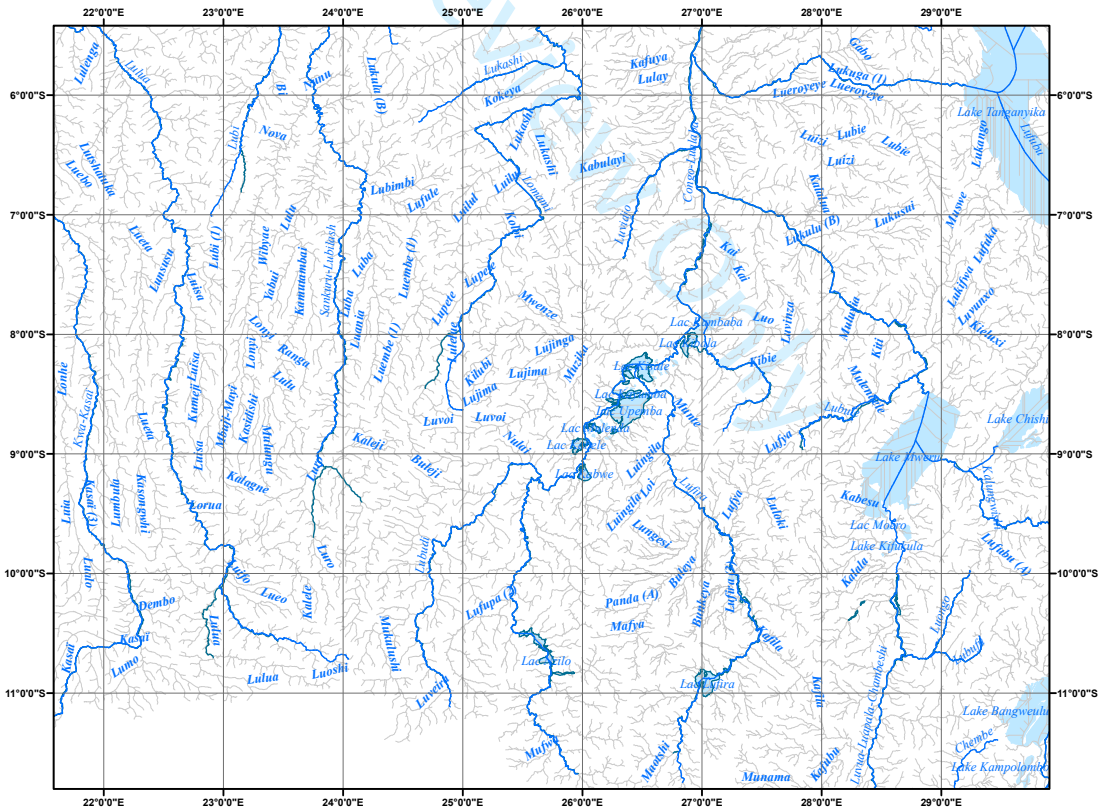


Figure 8 An excerpt of the identified river names within the CRB

### 3.2 Data availability and analysis

Increasing awareness of the value of the physical basin attributes that influence hydrological processes has driven efforts to develop useful approaches that can be used to identify and interpret regional patterns embedded in the observations. Such approaches consist of a variety of techniques for identifying groups of similar spatial objects and organizing them into hydrological typology. The usual approaches encompass the hierarchical and flat clustering algorithms, self-organizing maps, multivariate ordination, and hard versus soft classification, none of which are new to hydrological sciences (Nathan & McMahon, 1990; Olden *et al.*, 2011). However, these techniques use different algorithms for proximity measures and may yield different results when applied to the same dataset (Rao & Srinivas, 2006; Olden *et al.*, 2011). Olden *et al.* (2011) outline the deductive and inductive approaches for catchments similarity analysis. The deductive approach makes use of hydrologically relevant physical basin characteristics (e.g. climate, topography, vegetation, soils, and geology) that are assumed to control hydrological processes, in order to define the simple classification of contiguous or non-contiguous regions that are considered homogenous, with respect to certain environmental characteristics. This approach is useful when a general description of perceived hydrological patterns based on first principles is necessary to ease or advance understanding. The approach provides an alternative where observed streamflow data or reliably modelled hydrologic data are unavailable. The inductive approach uses signatures of the streamflow regime (e.g. magnitude, frequency, duration, timing, and the rate of change) to establish hydrologically similar groups. The inductive approach is the most reliable for catchment classification, though a challenge arises with the quality of the observed streamflow data and unsatisfactory records due to poor measurement. It should be noted that many of the classification signatures from Olden *et al.* (2011) have an ecological, rather than hydrological relevance; meaning that some these signatures may not link to hydrologic functions of the catchment. In this study, an attempt is made to use estimates (descriptors) of the physical basin attributes to build an a priori classification. Here “attempt” implies that there is no uniquely defined way in which the analysis can be performed and several approaches are explored.

#### 3.2.1 Estimates of the basin physical attributes

Various global datasets were assessed to identify physical basin attributes necessary to develop relationships between various features of the basin physiographic settings. The descriptors of the catchments characteristics used are based on the basin physical properties and consist of dimensional basin attributes as well as dimensionless indices. The dimensional attributes include climate input, topography, proportion of the basin area under vegetation cover types, proportion of the basin area under various soil particle sizes (clay, sand and silt). Blöschl (2005) mentions the important role of similarity indices which are usually expressed as dimensionless numbers

and reflect some invariant properties of the catchment functioning. The dimensionless similarity measures of the physical basin properties used in this study include the Ratio of long-term average evapotranspiration to the long-term average precipitation (PE/P: [Budyko, 1974](#)), a measure of the shape of the hypsometric form that is expressed by the hypsometric integral (HI: [Vivoni et al., 2008](#), [Deckers et al., 2010](#)), and the Topographic Wetness Index that reveals a topographic control on hydrological processes ([Sørensen et al., 2006](#)). Overall, 26 variables of the physical basin properties were derived from this analysis. Figure 9 shows a correlation matrix of the 26 variables assessed from the available datasets of the physical basin properties and that are considered to possess hydrological meaning. The analysis is made using XLSTAT (Addinsoft, 2020). Table 1 provides a description and source of the data used. Assessing correlations between the variables is important in order to examine patterns of variability among the basin characteristics and also to identify highly correlated variables in order to reduce the number for subsequent cluster analysis.

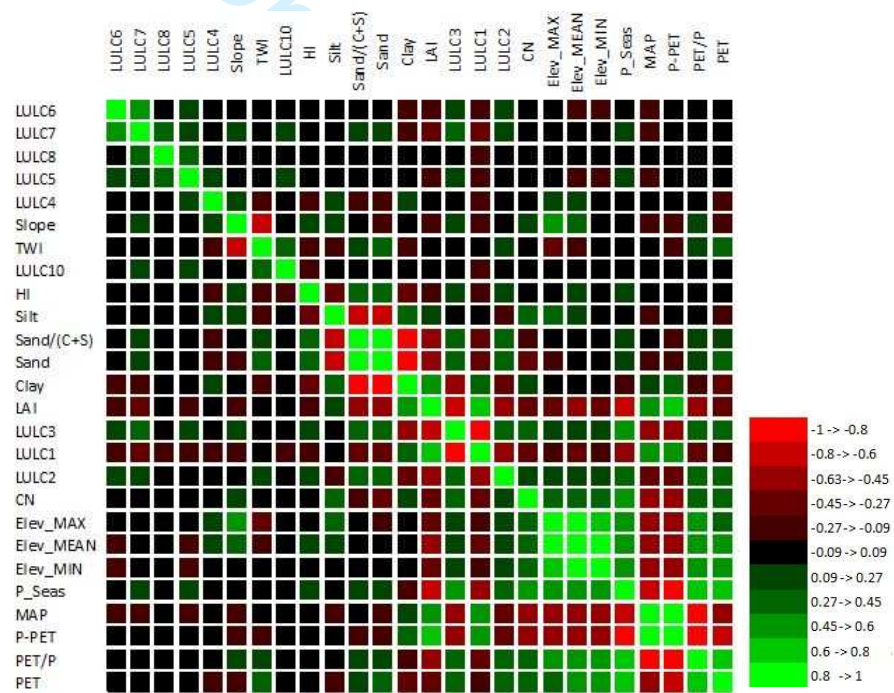


Figure 9 Pearson correlation matrix of the 26 variables of the physical basin properties. Refer to Table 1 for definition of the variables.

3.2.2 Ordination by Principal Component Analysis

The variables identified in this study are taken from samples of different scales. Furthermore, the correlation analysis (Figure 9) shows that there are variables that are highly correlated, thus suggesting that they are unlikely to contain any additional information and could be rejected to avoid redundant information. A Principal Component Analysis (PCA) was therefore conducted

with the aim to assign equal weight to all of the environmental variables included in the analysis, regardless of their scale of measurement so that the outcomes are not impacted by the effect of scales (Clarke & Warwick, 1994). PCA in this study also aimed to explain the variability of the environmental attributes by a small number of components as well as to reduce the dimensionality of a dataset consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the dataset. The result is a transformation of the original dataset into a new set of variables (the principal components-PCs), which are uncorrelated but ordered, and for which the first few PCs retain most of the variation present in all of the original variables (Jolliffe, 2002). The principal component and correlation analyses in this study were conducted using XLSTAT (Addinsoft, 2020). The measure of sample adequacy of Kaiser-Meyer-Olkin (KMO) was used to check the suitability of the samples for PCA. The KMO values range between 0 and 1, with a low value corresponding to the case where it is not possible to extract synthetic factors (or latent variables). Based on the KMO test, 15 variables were found to have KMO values greater than 0.5 and they were therefore selected for further analysis (Table 1). Figure 10 shows the ordination plots of the variables and their relationships with the catchments, after discrimination of the sample based on the KMO measure of sampling adequacy. More than 80 % of the variation is explained in the first five principal components (PCs), and the first two PCs explain 60% of the variation in the original variables. This is a good indicator of sampling adequacy for the PCA. PCA results show defined relationships between descriptors of the physical basin properties and the spatial distribution and grouping of the sub-basins along the principal component axes in a low-dimensional ordination space in which similar sub-basins are close together and dissimilar sites are far apart (Poff *et al.*, 2006).

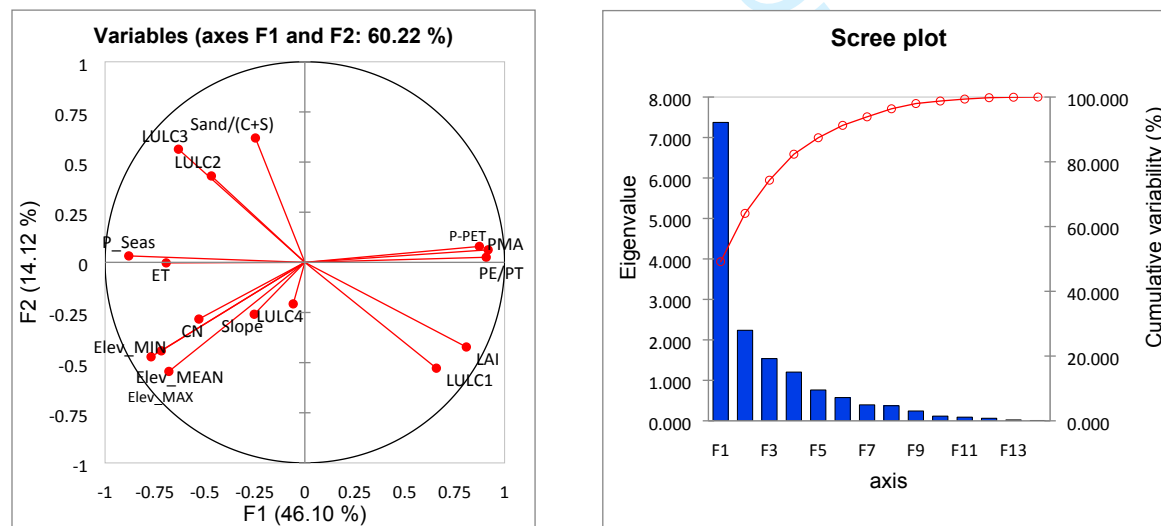


Figure 10 PCA ordination bi-plots for 15 variables physical basin attributes (left) and the scree plot of the Eigen value (right).



The first PCA axis (X axis) with an Eigen-value of 7.4 and the second PCA axis (Y axis) with an Eigen-value of 2.2 account for 46.1 % and 14.12 % of the total variation respectively. Therefore, the first two PCA axes explain 60.22 % of the variation in the variables of the physical basin properties. PC1 has a positive correlation with the variables of climate (MAP, PET/P, P-PET), tree cover area and LAI. These variables are likely to explain relationships of catchments with high rainfall and dense forest cover. A negative correlation is observed on this axis with the variables Elevation, precipitation seasonality, slope, PET, CN, LULC2, LULC3, and the fraction of soil particle size (Sand/(Clay+Silt)), which implies that these variables are very low in the catchment with high rainfall and dense forest cover. Based on these relationships, a cluster analysis was carried out to identify regions of homogeneous physical basin properties.

Figure 11 shows the correlation matrix based on the Pearson correlation coefficient with a p value of 0.05. In practice, the highly correlated variables are unlikely to contain any additional information and could be rejected to avoid redundant information, although Wagener *et al.* (2004) suggest using caution for correlation analysis and subsequent reduction of variables. In this study, the descriptors of elevation (Min, Max and Mean) and climate (MAP, PET/P, P-PET) show a strong correlation between them. However, the training of cluster analysis showed that these variables play a major role in partitioning some relevant features that have a meaning from the perceptual understanding of the basin physiographic setting. This is the case with a combination of the variables PET/P and P-ET that play a role in distinguishing relevant physiographic features that are characteristics of the central part of the Congo basin (Figure 19)

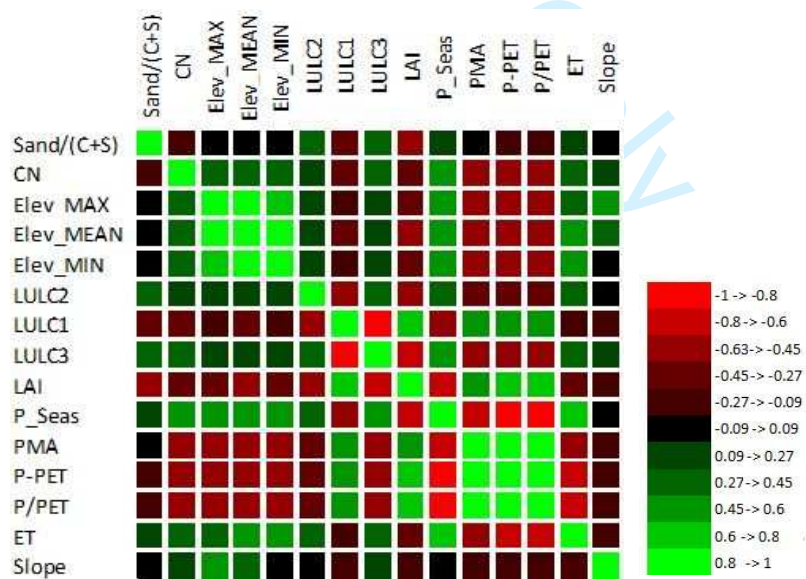


Figure 11 Correlation matrix (Pearson (n)) showing relationships between the 15 variables used in PCA.

Table 1 Statistical and sampling adequacy characteristics of the selected variables of the physical basin properties

No	Variable	Description	Unit	Obs.	KMO	Min	Max	Mean	Std. dev.
1	MAP	Long term mean annual precipitation (FAO Local climate)	[mm]	1740	0.7	798	2225	1561	274
2	PET	Long term mean annual evapotranspiration (FAO Local climate)	[mm]	1740	0.6	908	2292	1307	123
3	PET/P	Local rainfall deficiency: Ratio of the long-term PET and P (Budyko, 1974)	[-]	1740	0.7	0.47	2.7	0.9	0.24
4	P-PET	Available water	[mm]	1740	0.7	- 1429	1142	254	354
5	P_Seas	Precipitation Seasonality (Coefficient of Variation), Fick and Hijmans (2017)	[-]	1740	0.9	22	122	58	23
6	Sand/(C+S)	Soil fraction of sand, Clay and Silt	[-]	1740	0.6	0.3	3.1	1.0	0.5
7	CN	Curve Number (0.1°x 0.1°) utilizing the latest MODIS land cover and the Harmonized World Soil Database (Zeng et al. 2017)	[-]	1740	0.9	30	93	70	18
8	Slope	Derived from MERIT DEM based on a script that uses a filtering procedure (Hengl et al., 2003)	[%]	1740	0.6	0.1	14.8	2.6	1.9
9	Elev_Min	Minimum elevation derived from MERIT DEM	[m]	1740	0.8	-6.6	1631	548	259
10	Elev_Max	Maximum elevation derived from MERIT DEM	[m]	1740	0.8	264.7	4472.9	919.6	468
11	Elev_Min	Mean elevation derived from MERIT DEM	[m]	1740	0.8	51	2299	694	319
12	LAI	Leaf Area Index	[-]	1740	0.9	0.7	5.2	2.8	1.3
13	LULC1	Tree cover area (20 m resolution)	[%]	1740	0.8	0.2	100	70.7	27.5
14	LULC2	Shrub cover area (20 m resolution)	[%]	1740	0.8	0.0	77.5	2.9	6.2
15	LULC3	Grass land (20 m resolution)	[%]	1740	0.8	0.0	91.3	20.7	23

### 3.2.3 Hierarchical Agglomerative Cluster Analysis

Hierarchical clustering is an unsupervised classification method (Ley *et al.*, 2011) that consists of an iterative process by which either smaller clusters are combined into larger ones (agglomerative), or larger clusters are split into smaller ones (divisive) to produce a classification of objects typically presented as a dendrogram of clusters (Olden *et al.*, 2011). Rao & Srinivas (2006) give a list of several representative algorithms for Hierarchical Agglomerative Clustering (HAC). These algorithms differ in the way they compute the similarity between a pair of clusters and there seems to be no rules for the selection of a particular algorithm. In hydrology, the Euclidean distance appears to be the most frequently used method and has been applied in combination with different linkage algorithms such as the unweighted group average distance (e.g. Ley *et al.*, 2011), single and complete linkages (e.g. Tasker, 1982; Rao & Srinivas, 2006), group centroid (e.g. Pegg & Pierc, 2002), Ward's algorithm (e.g. Hosking & Wallis, 1997; Mazimavi, 2003; Rao & Srinivas, 2006). In the CRB, there are very few studies that assessed the application of the cluster analysis (Tshimanga, 2012). In the present study, a set of cluster analysis algorithms were tested, upon which the Chebychev distance in combination with the Ward's linkage algorithm was used. Chebyshev distance (or Tchebychev distance), is a metric defined on a vector space where the distance between two vectors is the greatest of their differences along any coordinate dimension (Cantrell, 2000). It is also called maximum value distance or  $L^\infty$  metric, and computes the absolute magnitude of the differences between coordinates of a pair of objects (equation 1).

$$d(x,y) = \max (|x_1 - x_2|, |y_1 - y_2|) \quad \text{Eq1}$$

In Ward's method (minimal increase of sum-of-squares), proximity between two clusters is the magnitude by which the summed square in their joint cluster will be greater than the combined summed square in these two clusters.

## 4. Results and discussion

The results of the a priori catchment classification approach presented for this study consist of a three-level partition of the landscape units, the unsupervised classification based on HAC analysis, and the multivariate statistics of the physiographic groupings.



## 4.1 Three-level partition of the landscape units

### 4.1.1 Catchment units along topographic gradients

Figure 12 shows the major landforms of the CRB and the resulting catchment units that translate the most frequent and dominant areas across the elevation and slope gradients. The elevation gradient determines the concentric landform layers for which the surface area decreases with the increasing elevation. The slope map is classified into seven groups where the slope class of 0-0.25% appears to define well most catchments with wetlands or natural lakes. In contrast of soil sciences where classes of soils have often been based on slope gradient, there appear to be no guidelines for the specific application of slope classes for hydrological purposes. The slope classes proposed by Nachtergaele (2010) indicate seven categories of slopes ranging from Flat wet (0-0.5%), Flat (0-2%), Undulating (2-8%), Rolling (8-15%), Moderately steep (15-30%), Steep (30-60%), Very steep (>60%). Engelen *et al.* (2006) use a classification that takes into consideration the landforms, ranging from Level land (<10%), Sloping land (10-30%) and Steep land (>30%). The slope class of 0-0.25% for flood plain and the standard slope of 5% are used for the Soil and Water Assessment Tool (SWAT, Neitsch *et al.*, 2009) and the Soil Conservation Service (SCS, Mishra & Singh, 2003), respectively. The strengths of the above-mentioned classifications were combined to derive a slope class map for the CRB as shown in Figure 12 (right).

The Cumulative Distribution Function (CDF) of the elevation distribution (normalized) is used to analyze the patterns of different geomorphological groups and forms associated with the corresponding catchment units. Similar to hypsometric curves, the shape of these curves provides valuable information on the dominant morphological features of the basin, hence the likely dominant hydrological processes (Vivoni *et al.*, 2008). In this regard, four major groups of geomorphological patterns result from the analysis, including the concave shaped curves (CC), convex shaped curves (CV), rectilinear shaped curves (R), and S shaped curves (S), which also all contain some internal variant forms (see Figures 13 and 14 that illustrate the patterns and spatial distribution of these geomorphological forms).

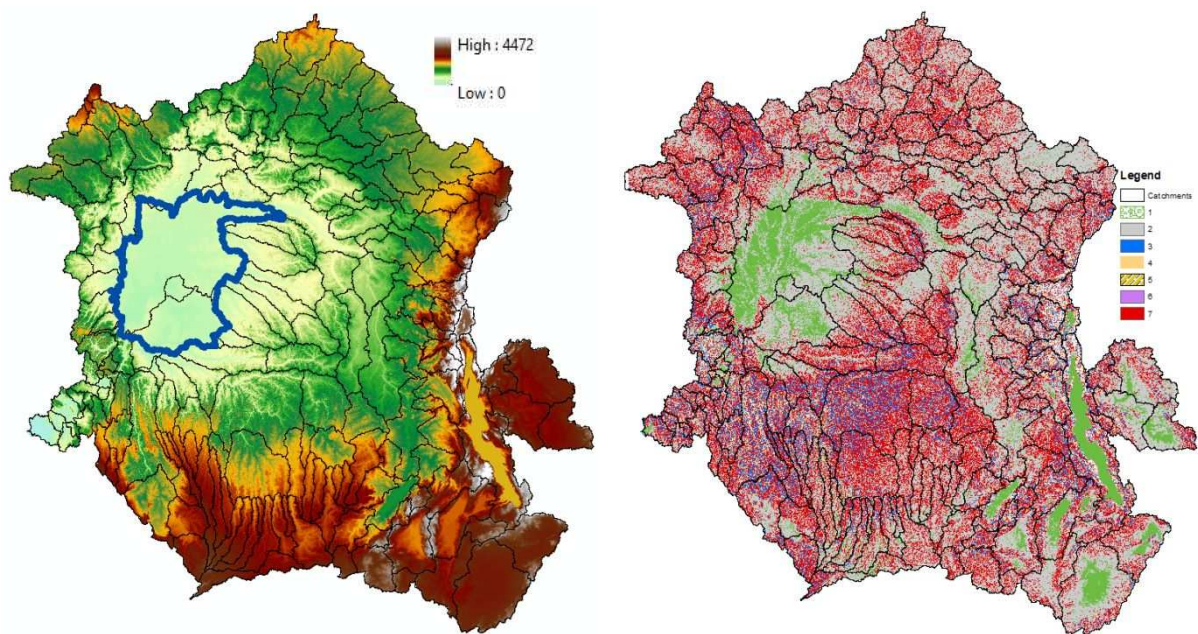


Figure 12 First level partition of catchment units along the elevation (left) and slope (right) gradients. The slope classes include 1 [0-0.25%], 2 [0.25-2%], 3 [2-5%], 4 [5-8%], 5 [8-15%], 6 [15-30%], 7 [30-84%].

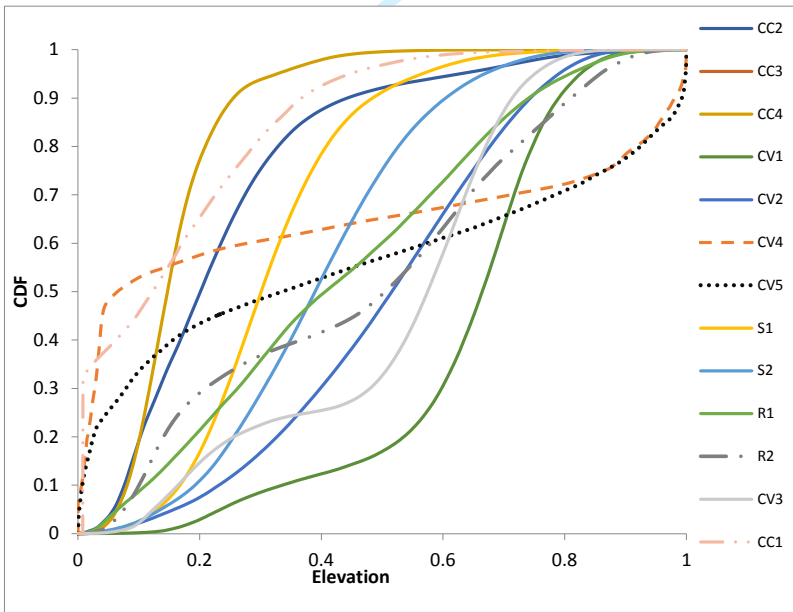


Figure 13 Cumulative Distribution Functions (CDF) forms across the dominant topographic gradients in the CRB [concave curves (CC), convex curves (CV), rectilinear curves (R), and S curves (S). Numbers refer to internal variant forms within the general shape].

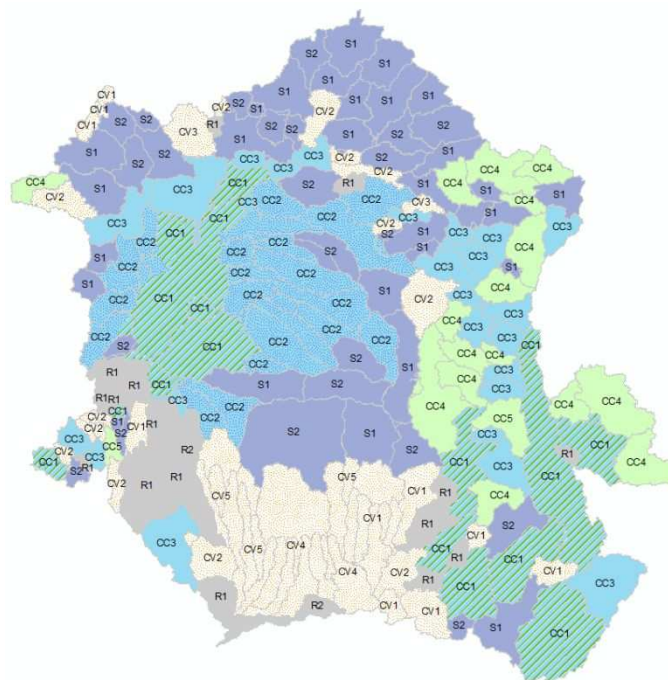


Figure 14 Patterns of different geomorphological groups and forms associated with the corresponding catchment units as derived from CDF. A single colour is used to represent sub-groups of the shapes S (S1, S2), R (R1, R2), CV (CV1,CV2,CV3,CV4, CV5)).

### ***Concave shaped curves***

The major features of the concave shaped curves that emerge from this analysis are those that mark the differentiation between differentiation of the wetland and natural lake catchments (CC1) from the other non-wetland catchments. The group of the catchments characterized by natural lakes and wetlands (Figure 15 presents the individual forms), include the lakes system of the Lualaba region: Bangweulu, Kamalondo depression, Kivu, Lufira reservoir, Moero and Tanganyika; the lake system of the Cuvette Centrale: May-Ndombe, Tele and Tumba; and the depression system of the Lower Congo: Pool Malebo and the littoral part of the basin. The Kwa River that constitutes a delta system of the Kasai River and Lake May-Ndombe is also represented in this group, and is known for its attenuation effect to the streamflow gauging site of Lediba (Tshimanga, 2012). These features are reflected in the shape of the CDF curves where the bottom part has a vertical line parallel to the Y-axis indicating the presence of Wetland or natural lakes. They are all dominated by lower land areas that are reflected in the concave shape of their curves as well. In the same group, there are some particularities related to the surface topography. For instance, a wetland that holds an island presents the horizontal line overlapping the X-axis. This is the case for catchment ID (CID here and after) 39 (Lake Kivu) located in the north of the Lake

Tanganyika and CID 111 (Pool Malebo) located in the most downstream of the Congo Basin in Kinshasa city.

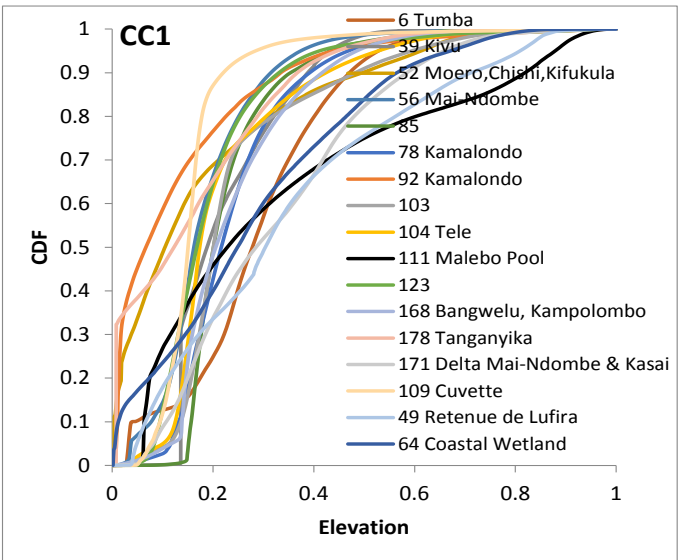


Figure 15 Variant forms of catchments characterized by concave shaped CDFs of the type of natural lakes and wetlands (refer to Figure 16 for the location of the feature ID)

On the other hand, the CID 178 which is the Lake Tanganyika does not hold an island and the length of its vertical bottom line as it is similar to CID 39 probably indicates the presence of high elevations surrounding the wetlands, given the fact that both catchments are located in mountainous region of rift valley. Alternatively, wetlands located in the most downstream of the CRB present a shorter vertical line. CID 111 (Pool Malebo) and CID 64 (Mangrove park at the mouth of the Congo River) illustrates this evidence. The central part of the cuvette central (CID 109) has more than 90 % of its total area occupied by low elevations, indicating that it accumulates more sediments than other sub basins of the same system. A close examination of these natural lake and wetland catchment based on Google Earth Engine analysis shows that they also take into consideration the areas covered by water during high flow seasons. Figure 16 shows a corresponding map of the wetland and natural lakes catchments.

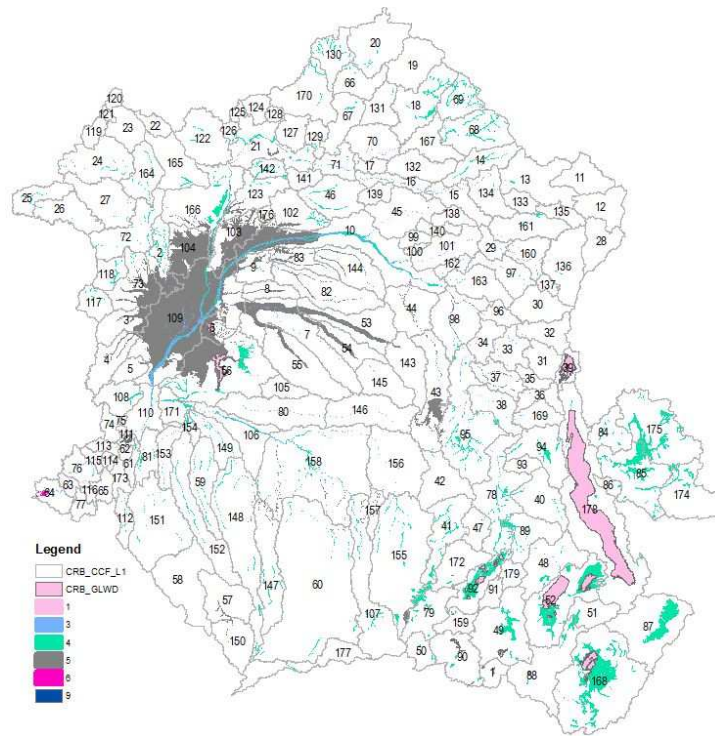


Figure 16 Map of the corresponding wetlands and natural lakes catchments [analysis done using 179 over 199 units]. **1:** Lake, **3:** Rivers, **4:** Freshwater Marsh/Floodplain, **5:** Flooded forest, **6:** Coastal wetland, **9:** Intermittent wetland/Lake.

The second variant in this category of concave shaped curves is related to the catchments with no wetlands (CC2, CC3, CC4). These catchments are represented by CDF curves showing a long upward concavity on their upper parts in an opposite seven number. They are dominated by a slope (2 – 8 %) falling in the class of undulating slope (Nachtergaele, 2010). These catchments present most of their basin area at lower elevations indicating a concave shaped catchment suggesting high erosion activity (Vivoni *et al.*, 2008). They have more than 90 % of their area dominated by lower elevations, which is likely to represent an advanced erosion process or sediment accumulation.

### **Convex shaped curves**

The second category of catchments has dominant area at higher elevations that are reflected through convex shape of CDF of the elevations at the bottom of the curves (CV1,CV2,CV3,CV4,CV5). Having a shape of this type would indicate that these catchments are at an infantile stage of the erosion process (Vivoni *et al.*, 2008) and the position of each curve situates the level to which the basin has been exposed to the erosion process. Furthermore, the convex shape of these catchments shades light on the likely dominant hydrological processes.



According to [Vivoni et al., \(2008\)](#), convex shapes have a high total runoff dominated by sub surface processes. Some distinctive variants of these forms are also illustrated below.

### ***S shaped curves***

These forms are a transition between concave and convex shaped curves, and are termed here as S shape curves (S1, S2). They dominate the northern catchments of the Oubangui and Sangha Rivers as well as they form the transition zone between the central part of the basin and the southern catchments in the Lunda highlands

### ***Rectilinear shaped curves***

The fourth category of the sub basins is represented by the CDF curve drawing 45° in a straight line like (R1, R2). They are stable sub-basins for the area of lower and higher elevations are almost balanced indicating a maturity stage.

#### **4.1.2 Socio-economic and anthropogenic systems units**

Figure 17 shows a spatial distribution of 203 catchments delineated based on the inclusion of anthropogenic features (urban areas, Mining areas, outlet of the historical streamflow monitoring stations and the reaches containing ports for river navigation), and that are nested within the first level partition of dominant morphological units (197), which all make a total number of 403 catchment units.

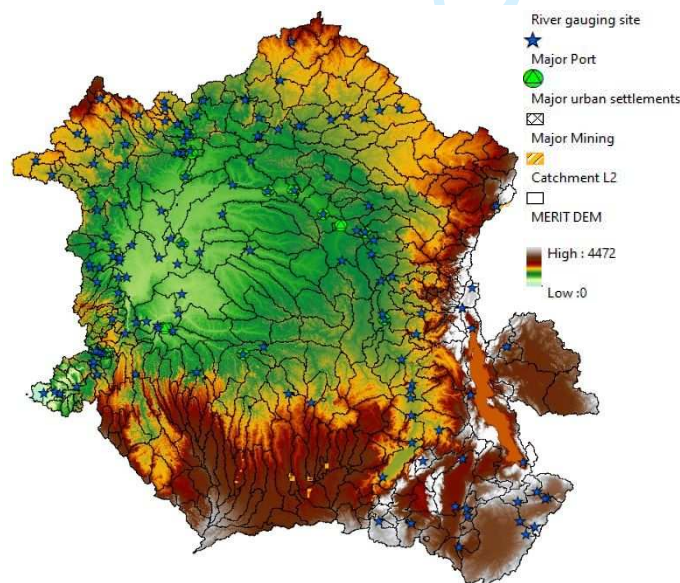


Figure 17 Spatial distribution of 403 catchment units including socio-economic and anthropogenic systems units

### 4.1.3 Water management units

Day to day issues of water management involve operational policies for water quality control and water supplies, flood monitoring, environmental water requirements, etc., all of which require information at the scale of decision making or catchment scale. Figure 18 shows the spatial distribution of 1337 catchments derived from the existing river names and are nested within the 403, which all makes a total number of 1 740 of units. The catchment area is a major parameter that is related to scale, Table 2 shows the statistics of the specific catchment areas derived in the study. The average catchment area obtained in this partition represents 2 118.5 km<sup>2</sup> with a Minimum area of 16.7 km<sup>2</sup> (an urban catchment of the Gombe River in Kinshasa) and a Maximum area of 70 317.7 km<sup>2</sup> (representing the Lake Tanganyika). Table 2 shows an excerpt attribute table for a sample of 30 catchments.

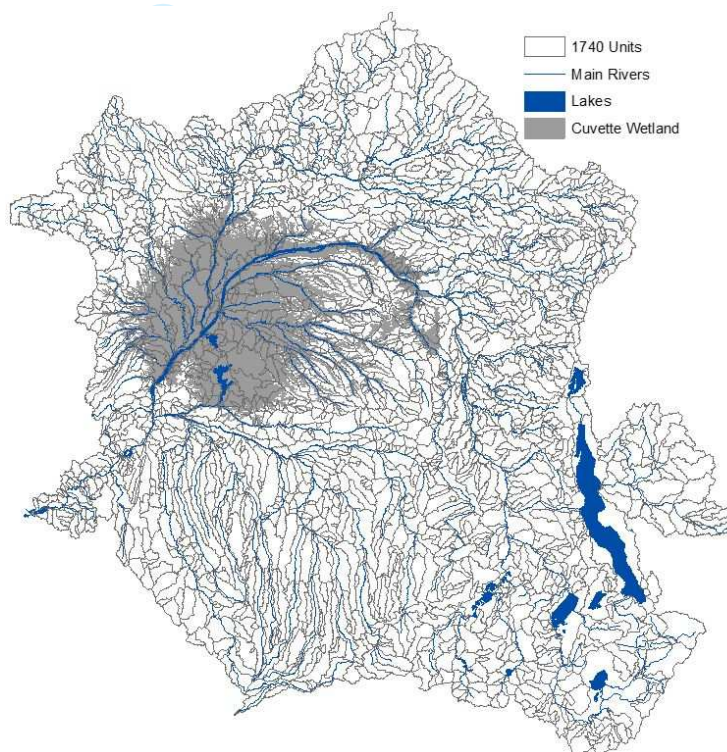


Figure 18 Spatial distribution of 1740 catchment units including the three-level partition



Table 2 An excerpt attribute table for 1740 units

Catchment ID	Area [km2]	Main River	Territory	Country
1	1284.9	Loya	Ituri	DRC
11	5762.6	Lukashi	Kabinda	DRC
21	124.4	Lulu	Tshopo	DRC
31	4927.2	Luele	Lunda Norte/Lunda Sul	Angola
41	1904.4	Mansya	Muchinga	Zambia
51	2700.8	Lukushi	Haut-Lomami/Tanganyika	DRC
61	4203.6	Lubudi (A)	Kolwezi	DRC
71	3053.6	Ruchuyi	Kigoma	Tanzania
81	4167.2	Lufuko	Tanganyika	DRC
91	519.2	Lubia	Tanganyika	DRC
101	577.6	Ambia	Haut-Uele	DRC
201	212.3	Peme	Bas-Uele	DRC
301	5736.1	Kotto	Haute-Kotto	CAR
401	416.3	Loloko	Sankuru	RDC
501	465.9	Mpulu	Kasai	RDC
601	1987.9	M'bali	Ombella-M'Poko	CAR
701	597.6	Mulendele	Haut-Katanga/Tanganyika	RDC
801	1825.3	Lokoli	Cuvette-Ouest	Congo
901	1069.6	Puluya	Maniema/Kabinda	DRC
1002	1236.0	Luvio	Sankuru	DRC
1101	534.6	Lubiri	Nord-Kivu	DRC
1201	1271.9	Lupoto	Haut-Katanga	DRC
1301	515.6	Lunaza	Haut-Katanga	DRC
1401	1583.8	Lukuga	Tanganyika	DRC
1501	2662.8	Congo a Inkisi	Cataractes/Lukaya/Pool/Kinshasa	DRC/Congo
1521	729.7	Komo	Likouala/Sangha	Congo
1522	12763.7	Likouala aux Herbes	Likouala/Cuvette	Congo
1601	491.8	Masso	Lobaye	CAR
1701	433.3	Musangazi	Central/Haut-Katanga	Zambia/DRC
1740	5008.7	Vovodo	Haut-Mbomou/Mbomou	CAR
Summary statistics of the total 1740 catchments				
Minimum	16.7	Gombe	Kinshasa	DRC
Maximum	70317.7	Lake Tanganyika	Tanganyika/Sud Kivu/Rukwa/Northern/Kigoma/Makamba	DRC/Burundi, Tanzania, Zambia
Sum	3686248.7	-	-	-
Median	1201.7	-	-	-
Mean	2118.5	-	-	-
Variance (n-1)	8542830.3	-	-	-
S.dev (n-1)	2922.8	-	-	-

## 4.2 Cluster analysis by Hierarchical Agglomerative Clustering

Many approaches used to partition a landscape into hydrological units exist, with their strengths and weaknesses (e.g. Neitsch *et al.*, 2009). Some of these approaches tend to generate an excessive number of catchment units for which estimating parameters for prediction may prove to be a very challenging task (Wagener *et al.*, 2004). The approach used in this study helped derive 1740 catchments that are representative of a range of climatic, geomorphologic, hydrologic, land use and water management domains in the CRB; thus, contributing to achieve one of the objectives of representing variability that is known to exist within the landscape units (Brown *et al.*, 2014). These catchment units were further used for analysis of catchment classification based on the unsupervised method of cluster analysis by the Hierarchical Agglomerative Clustering (HAC). A combination of the Chebychev algorithm and the Ward's linkage based on the 15 variables of the physical basin properties yielded 11 classes of physiographic homogenous groups. The Chebychev distance in combination with the Ward's linkage algorithm was found the most suitable in aggregating classification objects. Figures 19 and 20 show the graphical representation of the groups of homogenous catchments, based on the physical basin attributes as identified from the HAC. From the perceptual understanding of the physiographic setting of the CRB, it can be seen that the 11 physiographic groups yielded in this study based on the HAC capture the main physiographic regions of the CRB, including the partition of the central part of the basin into three major groups (Groups 8,9 and 10), the rift valley region (Group 3), the Mitumba highlands (Groups 1 and 4), the Katanga-Chambeshi region (Group 6), the Lunda Rise (Groups 5 and 7), and the Atlantic Rise (Group 11). Group 2 constitutes a transition region between the central basin and the highlands of the south, as well as the Asande Rise in the north. There are also some outliers that could be attributed to errors in the datasets used. Figure 21 presents a Parallel Coordinate Plot that shows the variability of the 15 variables of the physiographic descriptors within the 11 groups.

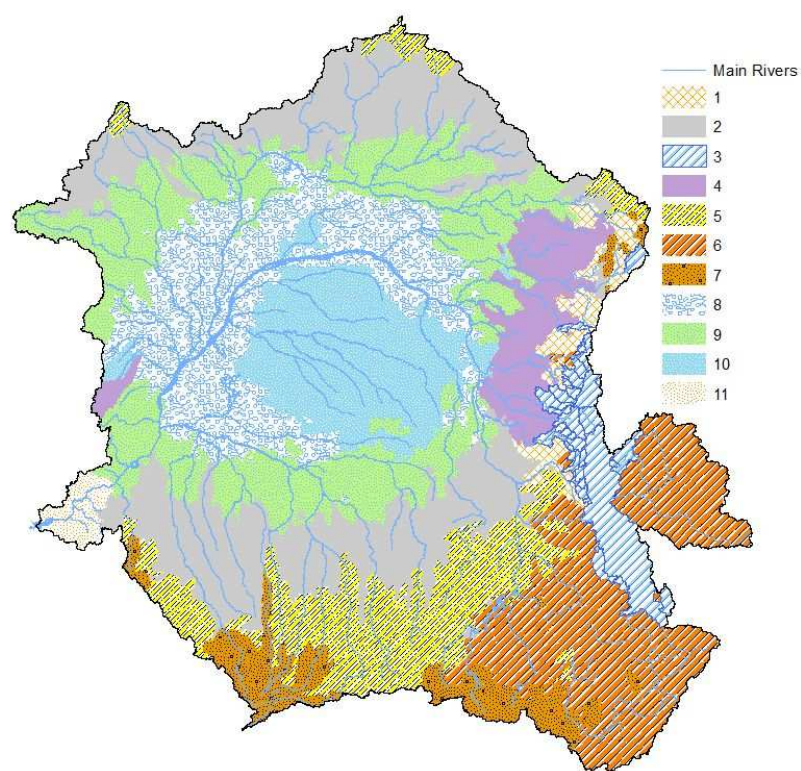


Figure 19      Graphical representation of 11 groups of homogenous physiographic setting as derived from the unsupervised classification based on HAC.

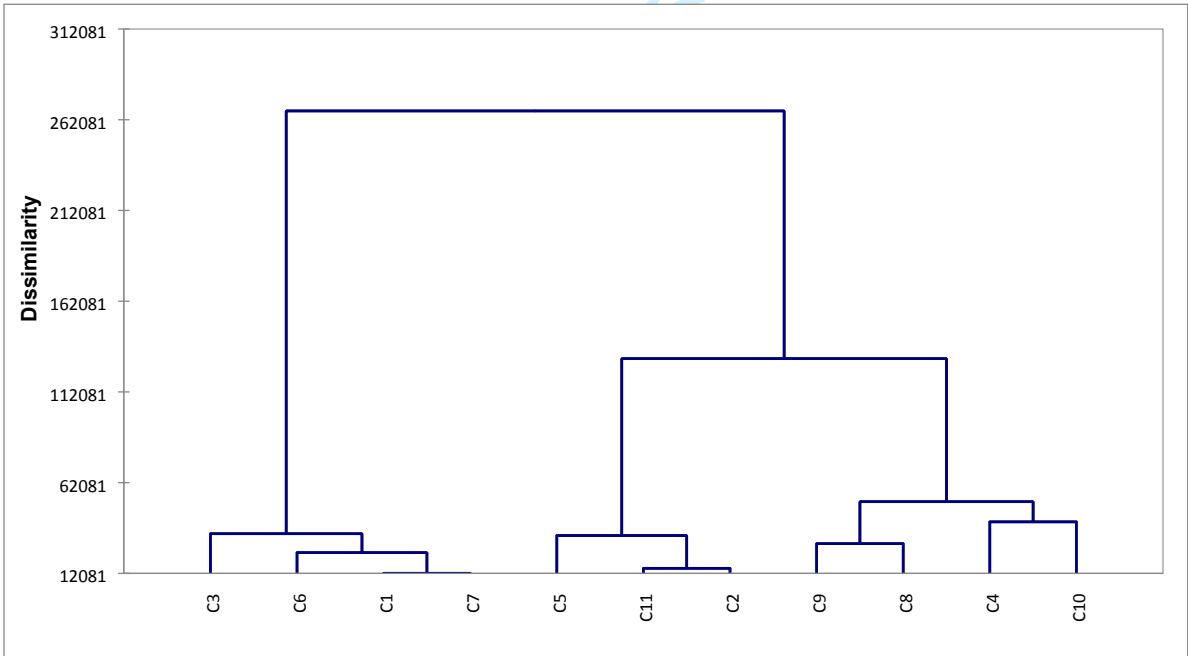


Figure 20      A dendrogram of the 11 groups of homogenous physiographic setting

The variance decomposition for the optimal classification shows 15.3 % of variance within classes and 84.7 % between classes. The similarity percentage analysis (SIMPER) is a more objective method of identifying similar and dissimilar characteristics between features and the approach is now widely used in environmental studies (Clarke, 1993).

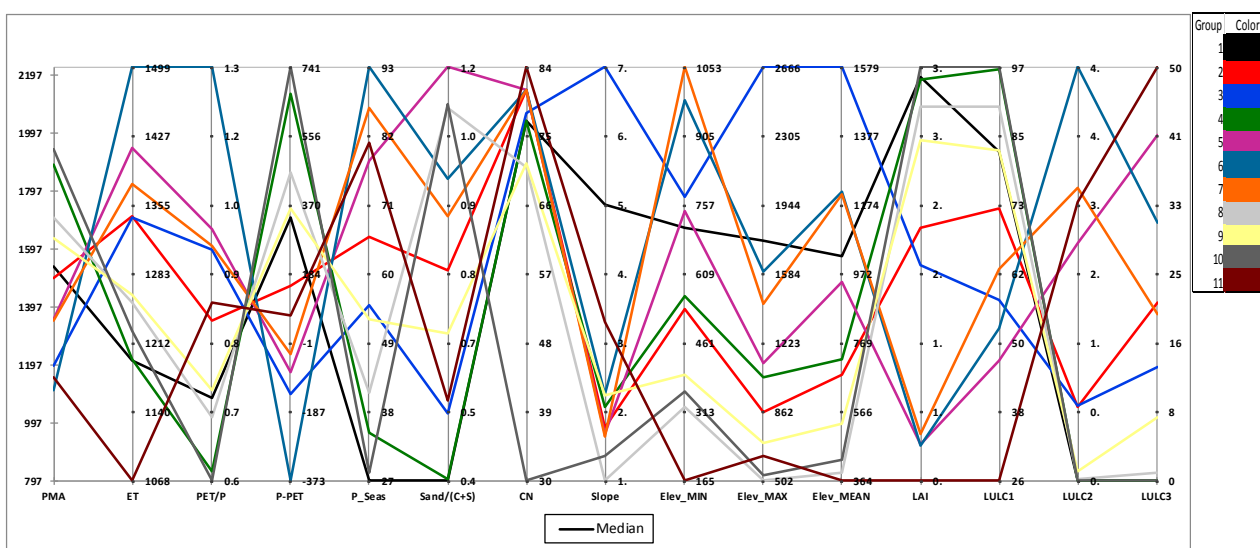


Figure 21 Parallel Coordinate Plot of the 11 groups of homogenous physiographic setting

In this study, SIMPER is used to identify variables responsible for similarity within the 11 regions of homogeneous physical basin attributes. The analysis was carried out using the PRIMER software package version 7. Prior to the SIMPER analysis, a global analysis of similarity (ANOSIM: Clarke & Warwick, 2001) test was conducted to assess the global characteristics (Global R) of the sample, where the overall R of 0.8253 with the p value of 0.0001 was obtained (Table 3), suggesting that there is a significant dissimilarity between the identified groups of physiographic properties.

Table 3            Matrix of R for the 11 groups identified in this study.

Group 1											
Group 2	0.88										
Group 3	0.69	1.00									
Group 4	0.79	0.80	0.99								
Group 5	0.84	0.63	0.99	0.94							
Group 6	0.87	0.99	0.81	1.00	0.74						
Group 7	0.83	0.91	0.92	0.97	0.51	0.37					
Group 8	1.00	0.91	1.00	0.96	1.00	1.00	1.00				
Group 9	0.99	0.54	1.00	0.86	0.99	1.00	1.00	0.50			
Group 10	1.00	0.97	1.00	0.87	1.00	1.00	1.00	0.64	0.84		
Group 11	0.98	0.91	1.00	0.99	0.99	1.00	1.00	1.00	0.97	1.00	

The results from the SIMPER analysis (Figure 22) show the overall contribution of the main variables to the average similarity within regions of homogeneous physiographic settings. The overall analysis of similarity shows that the topography and climate variables (Max elevation, P-PET, Mean elevation, MAP, Min elevation and PET) account for about 75% of the average similarity within the identified 11groups. Table 5 provides a summary description of the main characteristics of these 11 groups of physiographic setting for the CRB.

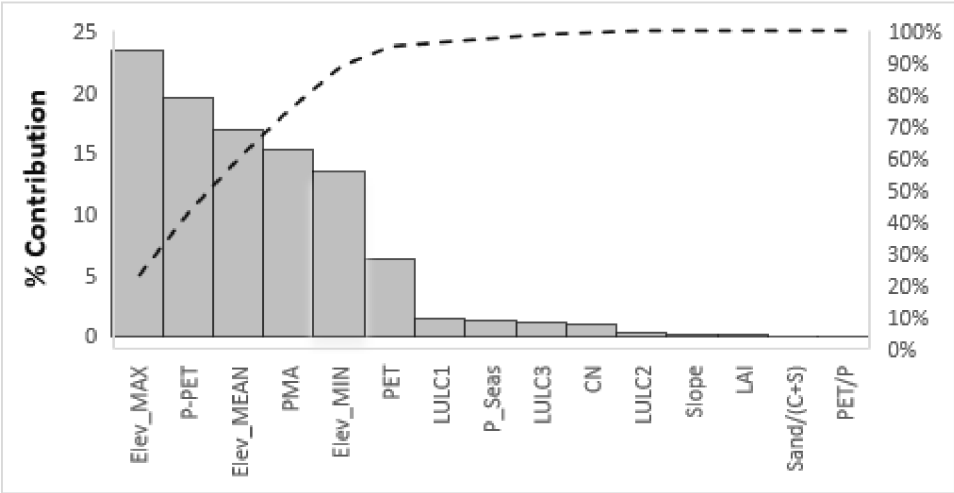


Figure 22            Overall contribution of the main variables to the average similarity within the regions of homogeneous physiographic settings.

Table 5 Summary description of the main characteristics of these 11 groups of physiographic setting for the CRB.

Group	Main characteristics	No of Units	% Sim.
1	The cluster occurs in the Mitumba highlands and constitutes the divides between the Congo and Nile River basins. Runge (2008) notes that the granite-dominated Mitumba mountain chain was uplifted to a height of over 3000 m. This higher local relief resulted in a high gradient and erosivity that changed the character and the influx of right bank tributaries to the Congo-Lualaba drainage system. It is characterized by high level of clay and silt particle sizes, high precipitation, and carries dense tropical rainforest.	35	2.1
2	The cluster dominates the main region of the northern part of the basin composed of the Oubangui River (Asande Rise) and it extends to the Cameroon highland where it drains the north eastern catchments of the Sangha River. This cluster is also found in the southern part of the CRB where it makes a transition belt between the central part of the basin and the southern highlands of Katanga, Angola (Lunda and Atlantic Rises). The Group partly drains semi-humid catchments with alternating dry and rainy seasons on both sides of the Equator (Runge, 2008). In the northern part of the Asande Rise, it carries wooded savanna, and in the south of it a mosaic of high woody biomass and patches of tropical forests in an area of 1600–1800 mm annual rainfall (Boulvert, 1996).	325	18.7
3	The cluster characterises the rift valley region including the lakes Tanganyika and Kivu, and the Ruzizi wetland. Pickford <i>et al.</i> (1993) note the evidence for an early Pliocene lake (Lake Obweruka) 550 km long, expanding in a north–south direction, which was subsequently separated into today’s smaller freshwater lakes of Kivu, Edward, and Albert. This cluster constitutes the source of the majority of tributaries that join the course of the Lualaba River from East, including Lukuga, Luama, Elila, Ulindi, Lugulu, Luka, Oso and Lowa Rivers, thus significantly contributing to an increase of the discharge of the Lualaba River before Kisangani.	31	1.8
4	Constitutes a transition zone between the Mitumba mountains and the Central part of the Congo basin and hold the majority of tributaries that join the upper Lualaba from the east, draining the western slopes of the Central African Rift valley. On average, rainfall in excess of 2000 mm falls annually on these slopes (Bultot, 1971). The rivers such as Elila, Ulindi, Lugulu, Lowa, Maiko, Lindi, Epulu, Nepoko and Bomakandi originate from this group.	111	6.4
5	The Group is characteristics of the Kasai Shield (Cahen and Snelling, 1966), it dominates the headwaters of Kasai and Lomami Rivers. The soil texture is dominated by sand fraction. Another characteristic of this group is the existence of parallel drainage lines with little convergence between them. The most northern part of this Group, characterizes the upper Kotto, with the presence of a Cretaceous carbonate formation which represents a complex structure of dual porosity, inter-granular porosity, joints and fractures, and local karstification.	145	8.3
6	Dominated by the Katanga-Chambeshi region, also known as the Paleo-Chambeshi drainage system, which is currently represented by the Upper Zambezi-Okavango and Lualaba-Luapula systems (Cotterill, 2005; Tshimanga, 2012). This system consists of the	158	9.1

	headwaters where the Congo River originates at an altitude of 1400–1500 m in the savanna highlands of the copper province, thus consisting of several small headwater streams, swamps and lakes (Runge, 2008). It also consists of the rivers draining the Chambeshi system through Lake Mweru, and Malagarasi system through Lake Tanganyika. The characteristics of this region are low slopes and low rainfall. The evapotranspiration is greater than the rainfall, which shows deficit in total annual availability. Sandy clay and loam for the topsoils, and clay loam for the sub-soil dominate the soil texture.		
7	This cluster constitutes the drainage divide between the Zambezi and Congo Rivers in the southern highlands of the CRB. This group is also characterized by the existence of parallel drainage lines with little convergence between them. The group constitutes the main source of the major rivers of the Kasai drainage area, including Kwango, Kasai, Chikapa (Tshikapa), Loange, Lushimo, Kwilu. This group is also found in the headwater of Lualaba where it contains the sources of the rivers such as Lubudi, Lualaba, Musoshi, Kafubu.	84	4.8
8	The cluster dominates the main wetland of the Cuvette Centrale (Bwangoyi et al. 2010) and is dominated by a slope class of 0-0.25%, which is characteristic of wetland. The major lakes of the Cuvette Centrale are found in this cluster, including Lakes Mayi-Ndombe, Tumba and Tele. Dargie et al. (2017) note the presence of extensive peat deposits that store below-ground 31 PgC of organic carbon. This cluster extends from the Aruwimi River to the confluence of the Kasai River, with the river bed characterized by anastomosing river channels, extended sand bars and is in places up to 16 km wide. Runge (2008) note that the shape of the approximately 1000 km wide central Congo Basin is round to oval, which earlier gave rise to the hypothesis of a former, enormous, so-called ‘Congo Lake’.	298	17.1
9	constitutes the transition zone between the lower and densely forested land of the Cuvette Centrale and high lands towards the northern part of the basin (Cameroonian highlands and the Bongo massif), southern part of the basin (Lunda plateau), and the eastern part of the basin (Mitumba mountain).	346	19.9
10	Dominates the drainage system of the Centrale region of the CRB, which is composed of the rivers such as Lopori, Maringa, Ikelemba, Tshuapa, Lomela, Busira, Salonga and Momboyo. The Mongala River on the right bank of the Congo main stem also exhibits the characteristics of this cluster. Extensive peat deposits store below-ground 31 PgC of organic carbon are found in this area (Dargie et al., 2017).	170	9.8
11	The cluster constitutes the littoral part of the basin where the Congo River enters the Atlantic Ocean and is part of the Atlantic Rise. The main tributaries that are generated in this region include Luozi, Lufu, Kwilu and Mponzo Rivers. The Coastal wetland “Parc Marin de Mangrove” is part of this cluster, which also has a very high hydropower potential (the Inga scheme) worth of more than 44 000 MW. Administratively, it includes the DRC’s territories of Luozi, Matadi, Songololo, Boma, Moanda and part of Mbanza Ngungu. According to Runge (2008) from Matadi to the Atlantic (134 km) the Congo River becomes part of an estuarine coast. The river’s bed is narrow and incised up to 250 m into outcropping quartzites. This section shows a winding pattern with huge whirlpools.	37	2.1
	<b>Total</b>	<b>1740</b>	<b>100</b>



## Conclusion and perspectives

The ability to predict the characteristics of hydrological dynamics is central to hydrology. This study consisted of establishing a framework of catchment classification based on available physical basin characteristics. Various approaches, including GIS and multivariate analyses, were combined to define the sub-basin units and group them in homogeneous regions. 1740 catchments were partitioned across the whole CRB using a multi-level criterion. Relevant data for the basin physiographic and climatic properties were used to generate the attribute values that were then explored further for similarity between different catchments. Information for sub-surface or geological characteristics was not readily accessible for the whole CRB and could not be used in this analysis to discriminate homogenous groups. Assessment of relationships between physical basin attributes and the catchments through PCA showed that 15 variables out of 26 initially selected contained the information necessary to explain the spatial distribution and grouping of the sub-basins. These 15 variables were included in a cluster analysis using unsupervised classification (hierarchical agglomerative clustering), which identified 11 groups of homogenous catchments based on the use of the environmental descriptors derived from the available physical basin property datasets. As far as prior knowledge of the basin physical processes is concerned (perceptual understanding), the identified homogenous groups indicate the main regions where there are similar patterns of environmental characteristics such as rainfall, evapotranspiration, topography, vegetation and soil types. This constitutes therefore an a priori clustering of the catchment classification framework for the CRB and should be considered as an initial phase of the iterative process of catchment classification. The question of whether the classes produced through this analysis are useful for the applications described in the introduction has not yet been addressed. Further work is needed to develop functional relationships of the catchment structures and processes necessary to address a range of needs for hydrologic prediction and water resources management in the CRB (refer to Figure 2). Additional information from signatures will help to evolve and refine the a priori classification.

## Acknowledgement

This research work is supported through the Royal Society Future Leader African Independent Research Programme - FLAIR, Grant: FLR\R1\192057.

## References

- Addinsoft (2020). XLSTAT statistical and data analysis solution. Long Island, NY, USA. <https://www.xlstat.com>.
- Beven, K., Asadullah, A., Bates, P., Blyth, E., Chappell, N., Child, S., Cloke, H., Dadson, S., Everard, N., Fowler, H. J., Freer, J., Hannah, D. M., Heppell, K., Holden, J., Lamb, R., Lewis, H., Morgan, G., Parry, L. & Wagener, T., 2019. Developing observational methods to drive future hydrological science: Can we make a start as a community? *Processes*. 2020;34:868–873. <https://doi.org/10.1002/hyp.13622>
- Blöschl, G., 2005. Statistical upscaling and downscaling in hydrology. *Encyclopaedia of Hydrological Sciences*. (Ed) Anderson. G.M., 2005. John Wiley & Sons, Ltd.
- Blöschl, G., Sivapalan, M., 1995. Scale issues in hydrological modelling: a review. *Hydrol. Process.*, 9, 251-290.
- Borges, A. V., Darchambeau, F., Lambert, T., Morana, C., Allen, G., Tambwe, E., ... Bouillon, S. (2019). Variations of dissolved greenhouse gases (CO<sub>2</sub>, CH<sub>4</sub>, N<sub>2</sub>O) in the Congo River network overwhelmingly driven by fluvial-wetland connectivity. *Biogeosciences Discussions*, in review. <https://doi.org/10.5194/bg-2019-68>
- Brown, S.C., Lester, R.E., Versace, V.L., Fawcett, J., Laurenson, L., 2014. Hydrologic Landscape Regionalisation Using Deductive Classification and Random Forests. *PLoS ONE* 9(11): e112856. doi:10.1371/journal.pone.0112856.
- Budyko, M. I., 1974. *Climate and Life*, 508 pp., Academic Press, New York
- Bull, L.J., Kirkby, M.J., Shannon, J., Dunsford, H.D., 2003. Predicting hydrologically similar surfaces (HYSS) in semi-arid environments. *Environ. Model. Software.*, 1 (2).
- Bultot, F., 1971. *Atlas climatique du bassin congolais*. Publications de l'Institut National pour l'Etude Agronomique du Congo (I.N.E.A.C., République Démocratique du Congo), 1–4.
- Boulvert, Y., 1996. *Etude géomorphologique de la République Centrafricaine : carte à 1/ 1000000 en deux feuilles Ouest et Est*. Paris : ORSTOM, (110), 258 p.
- Bwangoy, J.R.B., Hansen, M.C., Roy, D.P., Grandi, G.D., Justice, C.O., 2010. Wetland mapping in the Congo basin using optical and radar remotely sensed data and derived topographical indices. *Rem. Sens. Environ.*, 114 (1), 73-86.
- CAMI-Cadastre Minier, 2012. Concessions Minières en RD Congo. <http://congominer.org/reports/445-cartographie-cadastre-minier-rdc>
- Cantrell, C.D., 2000. *Modern mathematical methods for physicists and engineers*, Cambridge University press, London, pp 37–43
- Carr, A.B., Trigg, M.A., Tshimanga, R.M., Borman, J.D., Smith, M.W. 2019. Greater Water Surface Variability Revealed by New Congo River Field Data: Implications for Satellite Altimetry Measurements of Large Rivers. *Geophysical Research Letters* <https://doi.org/10.1029/2019GL083720>

- Carrillo, G., Troch, P.A., Sivapalan, M., Wagener, T., Harman, C., Sawicz, K., 2011. Catchment classification: hydrological analysis of catchment behavior through process-based modeling along a climate gradient. *Hydrology and Earth System Sciences*, 15, 3411–3430.
- Clarke, K.R., Warwick, R.M., 1994, Change in marine communities: an approach to statistical analysis and interpretation. Plymouth marine laboratory, UK.
- Crowley, J. W., Mitrovica, X.J., Bailey, C.R., Tamisiea, E.M., Davis, L.J., 2006. Land water storage within the Congo Basin inferred from GRACE satellite gravity data, *Geophys. Res. Lett.*, 33, L19402, doi: 10.1029/2006GL027070.
- Dargie, G.C., Lewis, S.L., Lawson, I.T., Mitchard, E.T., Page, S.E., Bocko, Y.E., Ifo, S.A., 2017. Age, extent and carbon storage of the central Congo Basin peatland complex. *Nature* 542, 86–90.
- Deckers, H.E.L.D., Booij, J.M., Krol, S.M., 2010. Catchment variability and parameter estimation in multi-objective regionalisation of rainfall-runoff model. *Water Resour. Manage*, 24, 3961–3985.
- De Smith, M.J., Goodchild, M.F., Longley, P., 2018. *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools*. 6th Edition, 2018. Matador: Leicester, UK, 2007; ISBN-13 978-1-912556-05-2.
- Fick, S. E., Hijmans, R. J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37, 4302–4315 (2017).
- Garbrecht, J., Martz, W.L., 1999. Digital elevation model issues in water resources modeling. In *Hydrologic and Hydraulic Modeling Support in GIS*, D. Maidment and D. Djokic, Eds., ESRI Press, Redlands, CA, 216 pp.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* 202, 18–27.
- Harvey, C.A., Gonzalez Villalobos, J.A., 2007. Agroforestry systems conserve species-rich but modified assemblages of tropical birds and bats. *Biodiversity Conservation*, 16, 2257–2292.
- Hosking, J.R.M., Wallis, J.R., 1997. *Regional frequency analysis: an approach based on L-moments*. Cambridge University Press, New York, USA.
- Hovius, N., Stark, C.P., Chu, H.T., 2000. Supply and removal of sediment in a landslide–dominated mountain belt: Central Range, Taiwan. *Journal of Geology*, 108, 73–89.
- Hughes, D.A., Tshimanga, R.M., Sithabile, T., Tanner, J., 2013. Simulating Wetland Impacts on Streamflow in Southern Africa Using a Monthly Hydrological Model. *Journal of Hydrological Processes* DOI: 10.1002/hyp.9725
- Jolliffe, I.T., 2002. *Principal Component Analysis*, Second Edition. Springer, New York.
- IRB (Institut Royal Belge), 1950. Canevas Géographique et Hydrographie, Musée Royal d’Afrique Centrale de Tervuren. <https://www.africamuseum.be/fr/home>

- Kadima, E., Delvaux, D., Sebagenzi, S.N., Tack, L., Kabeya, M., 2011. Structure and geological history of the Congo Basin: An integrated interpretation of gravity, magnetic and reflection seismic data. *Basin Research*, 23 (5), 499-527.
- Lexartza-Artza, I., Wainwright, J., 2009. Hydrological connectivity: Linking concepts with practical implications. Catchment Science Centre, Department of Geography, University of Sheffield, Winter Street, Sheffield, S10 2TN, UK.
- Laraque, A., Bricquet, J.P., Pandi, A., Olivry, J.C., 2009. A review of material transport by the Congo river and its tributaries. *Hydrological Processes* 23, 3216-3224.
- Laraque, A., Pouyaud, B., Rocchia, R., Robin, R., Chaffaut, I., Moutsambote, J.M., Maziezoula, B., Censier, C., Albouy, Y., Elenga, H., Etcheber, H., Delaune, M., Sondag, F., Gasse, F., 1998. Origin and function of a closed depression in equatorial humid zones: the lake Tele in north Congo. *Journal of Hydrology* 207, 236-253.
- Ley, R., Casper, M. C., Hellebrand, H., and Merz, R.: Catchment classification by runoff behaviour with self-organizing maps (SOM), *Hydrol. Earth Syst. Sci.*, 15, 2947-2962, <https://doi.org/10.5194/hess-15-2947-2011>, 2011.
- Maher, A., 1994. Africa-Europe Electrical Interconnection and Prospects of Worldwide Interconnections. Final Proc. Int. Conf. CIGRE Keynote Address Paris. 1994.
- Mayorga, E., Logsdon, M.G., Ballester, M.V.R., Richey, J.E., 2005. Estimating cell-to-cell surface drainage paths from digital networks, with an application to the Amazon basin. *J Hydrol* 315:167-182.
- Mazvimavi, D., 2003. Estimation of flow characteristics of ungauged catchments., Unpublished PhD thesis, Wageningen University and International Institute for Geo-Information and Earth Observation, ITC, Enschede, The Netherlands.
- Mukheibir, P., 2007. Possible climate change impacts on large hydroelectricity schemes in Southern Africa. *J. Energy Southern Africa.*, 18 (1), 4-9.
- Munzimi, Y. A., Hansen, M., Asante, O. K., 2019. Estimating daily streamflow in the Congo Basin using satellite-derived data and a semi-distributed hydrological model. *Hydrological Sciences Journal*, 64:12, 1472-1487, DOI: 10.1080/02626667.2019.1647342.
- Nachtergaele, A.F., van Velthuisen, B.H., Verelst, B.L., Batjes, C.N., Dijkshoorn, C.K., Vincent van Engelen, C.V., Fischer, B.G, Jones, D.A, Montanarella, D.L, Petri. A.M., Prieler, B.S., Shi, E.X, Teixeira, D.E., Wiberg, D.D., 2010. World Congress of Soil Science, Soil Solutions for a Changing World 1-6 August 2010, Brisbane, Australia. Published on DVD.
- Nathan, R.J., McMahon, T.A., 1990. Identification of homogeneous regions for the purposes of regionalisation. *J. Hydrol.*, 121, 217-238.
- Neitsch, L.S., Arnold, G.J., Kiniry, R.J., Williams, R.J., 2009. Soil and Water Assessment Tools Theoretical Documentation. Grassland, soil and water research laboratory. Agriculture research service. Texas 76502.

- Olden, J. D., Kennard, M.J., Pusey, B.J., 2011. A framework for hydrologic classification with a review of methodologies and applications in ecohydrology. *Ecohydrology* (wileyonlinelibrary.com) DOI: 10.1002/eco.251.
- Park, S., Kim, J., 2019. Landslide susceptibility mapping based on random forest and boosted regression tree models, and a comparison of their performance. *Appl. Sci.* 2019, 9, 942.
- Partow, H., 2011. Water issues in the Democratic Republic of the Congo. Challenges and opportunities. United Nations Environment Programme (UNEP), Technical report. Available from: [http://postconflict.unep.ch/publications/UNEP\\_DRC\\_water.pdf](http://postconflict.unep.ch/publications/UNEP_DRC_water.pdf)
- Pegg, M.A., Pierce, L.C., 2002. Classification of reaches in the Missouri and lower Yellowstone rivers based on flow characteristics. *River Research and Applications*, 18, 31–42.
- Pickford, M., Senut, B., Hadoto, D.P.M., 1993. Geology and Palaeobiology of the Albertine Rift Valley, Uganda-Zaire: Geology. *Geol.* 24, 1–1989.
- Poehls, J.D., Smith, J. G., 2009. *Encyclopedic Dictionary of Hydrogeology*. Academic press.
- Poff, N., Olden, J. D., Pepin, D. M., Bledsoe, B. P., 2006. Placing global stream flow variability in geographic and geomorphic contexts, *River Res. App.*, 22, 149–166.
- Rao, A.R., Srinivas, V.V., 2006. Regionalisation of watersheds by hybrid-cluster analysis. *J. Hydrol.*, 318, 37–56.
- Runge, J., 2008. The Congo River, Central Africa. In: Gupta, A. (Ed) *Large Rivers: Geomorphology and Management*. Wiley and Sons, London, UK.
- Sivakumar, B. and Singh, V. P., 2012. Hydrologic system complexity and nonlinear dynamic concepts for a catchment classification framework, *Hydrol. Earth Syst. Sci.*, 16, 4119–4131, <https://doi.org/10.5194/hess-16-4119-2012>, 2012.
- Sivakumar, B., Singh, V.P., Berndtsson, R., Khan, S., 2015. Catchment classification framework in hydrology: challenges and directions', *Journal of Hydrologic Engineering*, vol. 20, no. 1, pp. A4014002. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000837](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000837)
- Sivapalan, M., 2005. Pattern, process and function: elements of a unified theory of hydrology at the catchment scale. In Anderson, M.G. (ed.) *Encyclopedia of Hydrological Sciences*. John Wiley & Sons Ltd.
- Sivapalan, M., Takeuchi, K., Franks, S.W., Gupta, V.K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J.J., Mendiola, E.M., O'Connell, P.E., Oki, T., Pomery, J.W., Schertzer, D., Uhlenbrook, S., Zehe, E., 2003. IAHS Decade on Prediction in Ungauged Basins (PUB), Shaping an exciting future for the hydrological sciences. *Hydrol. Sci. J.*, 48 (6), 857–880.
- Sørensen, R., Zinko, U., Seibert, J., 2006. On the calculation of the topographic wetness index: evaluation of different methods based on field observations. *Hydrol. Earth Syst. Sci. Discuss.*, 10, 101–112.
- Strahler, A.N., 1954. Statistical Analysis in Geomorphic Research. *The Journal of Geology* 62, no. 1 (1954): 1–25. <https://doi.org/10.1086/626131>.



- Tasker, G.D., 1982. Comparing methods of hydrologic regionalisation. *Water Resour. Bull.*, 18: 965-970.
- Thiem, M.L., Abel, R., Burgess, N., 2005. *Freshwater ecoregions of Africa and Madagascar: a conservation assessment*. Washington DC, USA: Island Press.
- Tshimanga, R.M., 2012. Hydrological uncertainty analysis and scenario-based streamflow modelling for the Congo River Basin. *PhD thesis, Rhodes University repository. South Africa*.
- Tshimanga, R.M., Hughes, D.A., 2012. Climate change and impacts on the hydrology of the Congo Basin: the case of the northern sub-basins of the Oubangui and Sangha Rivers. *Physics and Chemistry of the Earth* 50–52 (2012) 72–83
- Tshimanga, R.M., Hughes, D.A., 2014. Basin-scale performance of a semi distributed rainfall-runoff model for hydrological predictions and water resources assessment of large rivers: The Congo River, *Water Resour. Res.*, 50, doi: 10.1002/2013WR014310.
- Tshimanga, R.M., Hughes, D.A., Kapangaziwiri, E., 2011. Initial calibration of a semi-distributed rainfall-runoff model for the Congo River Basin. *Physics and Chemistry of the Earth* 36: 14-15.
- van Oost, K., Govers, G., Desmet, P., 2000. Evaluating the effects of changes in landscape structure on soil erosion by water and tillage. *Landscape Ecology* 15 (6), 577–589.
- Vivoni, E.R., Benedetto, D. F., Grimaldi, S., Eltahir, B.A.E., 2008. Hypsometric control on surface and subsurface runoff, *Water Resour. Res.*, 44, W12502.
- Vogel, R.M., Kroll, C.N., 1996. Estimation of baseflow recession constants. *Water Resources Management* 10, 303-320.
- Wagener, T., Sivapalan, M., Troch, P., Woods, R., 2007. Catchment classification and hydrologic similarity. *Geography Compass*, 1, 10.1111/j.1749-8198.2007.00039.x.
- Wagener, T., Sivapalan, M., McGlynn, B., 2008. Catchment classification and services –Toward a new paradigm for catchment hydrology driven by societal needs. In Anderson, M.G. (ed.) *Encyclopedia of Hydrological Sciences*. John Wiley & Sons Ltd.
- Wagener, T., Wheater, H.S., Gupta, H.V., 2004. *Rainfall-runoff modelling in gauged and ungauged catchments*. Imperial College Press, London, UK, 332pp. ISBN 1-860944-66-3.
- Ward, J.V., Tockner, K., Arscott, D.B., Claret, C., 2002. Riverine landscape diversity. *Freshwater Biology*, 47 (4), 517–539.
- Wigmosta, M., Prasad, R., 2005. Upscaling and Downscaling-Dynamic Models. *Encyclopaedia of Hydrological Sciences*. (Ed) Anderson. G.M., 2005. John Wiley & Sons, Ltd.
- Woods, R., 2005. Hydrologic Concepts of Variability and Scale. In Anderson, M.G. (ed.) *Encyclopedia of Hydrological Sciences*. John Wiley & Sons Ltd.
- Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O’Loughlin, F., Neal, J. C., Bates, P., 2017. A high-accuracy map of global terrain elevations. *Geophysical Research Letters*, 44(11), 5844–5853. <https://doi.org/10.1002/2017GL072874>.

Yuan, T., Lee, H., Jung, H.C., Aierken, A., Beighley, E., Alsdorf, D., Tshimanga, R.M., Kim, D. 2017. Absolute water storages in the Congo River floodplains from integration of InSAR and satellite radar altimetry. *Remote Sensing of Environment* 201 (2017) 57–72. <http://dx.doi.org/10.1016/j.rse.2017.09.003>.

For Review Only

## Towards a framework of catchment classification for hydrologic predictions and water resources management in the ungauged basin of the Congo River: An a priori approach

Raphael M. Tshimanga<sup>1</sup>, Gode Bola<sup>1</sup>, Pierre Kabuya<sup>1,2</sup>, Landry Nkaba<sup>1</sup>, Jeffery Neal<sup>3</sup>, Laurence Hawker<sup>3</sup>, Mark A. Trigg<sup>4</sup>, Paul Bates<sup>3</sup>, Denis A. Hughes<sup>2</sup>, Alain Laraque<sup>5</sup>, Ross Woods<sup>6</sup>, Thorsten Wagener<sup>6</sup>

1. Congo Basin Water Resources Research Center (CRREBaC) & Dpt. Natural Resources Management, University of Kinshasa, Box 117 KIN XI, DR Congo
2. Institute for Water Research, Rhodes University, Box 94, Grahamstown 6140, South Africa
3. School of Civil Engineering, University of Leeds, Leeds LS2 9JT, UK
4. School of Geographical Sciences, University of Bristol, UK
5. GET-UMR CNRS/IRD/UPS – UMR 5562 du CNRS, UR 234 de l'IRD – 14, av. E. Belin – 31400 Toulouse (France)
6. Department of Civil Engineering, University of Bristol, UK

[raphtm@yahoo.fr](mailto:raphtm@yahoo.fr)

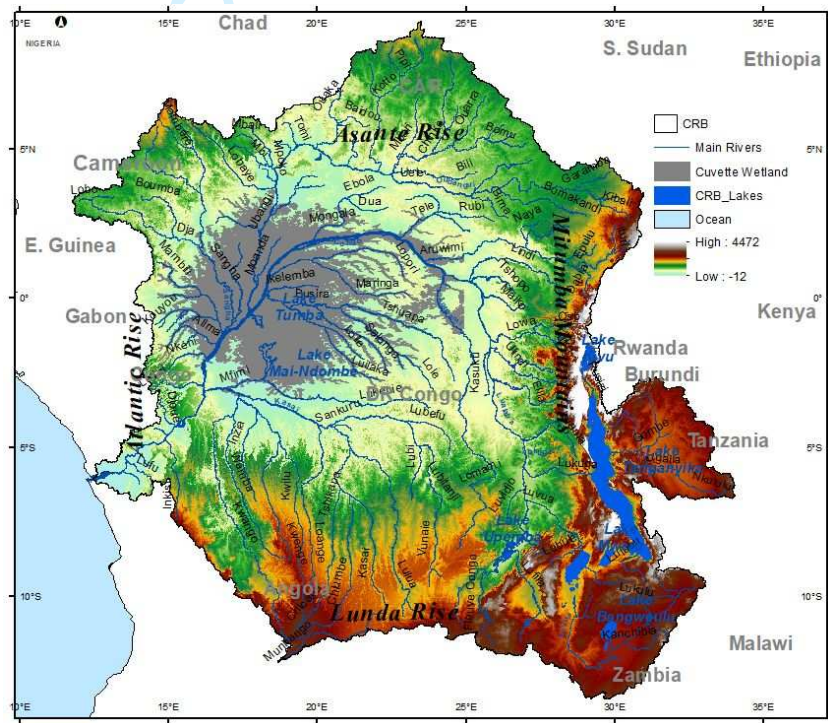


Figure 1 Physiographic setting of the Congo River Basin

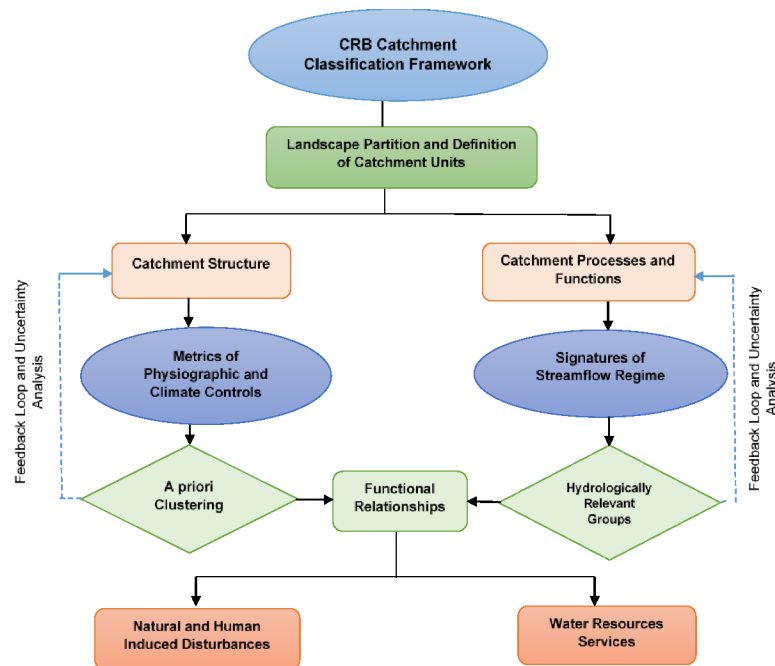


Figure 2 A conceptual framework of catchment classification for the CRB

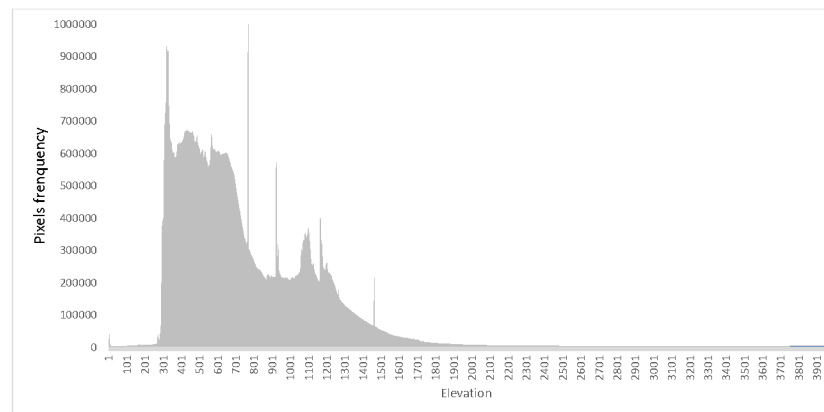
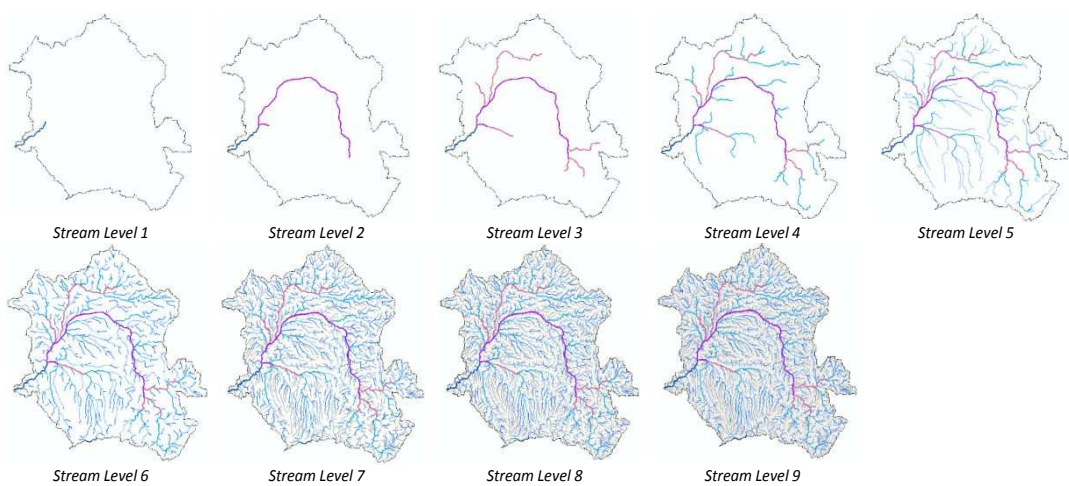
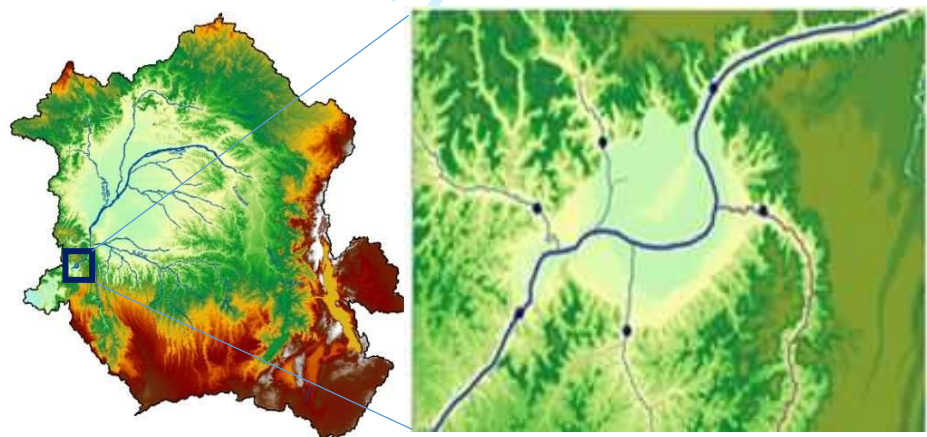


Figure 3 Frequency distribution of the elevation structure in the CRB from MERIT DEM

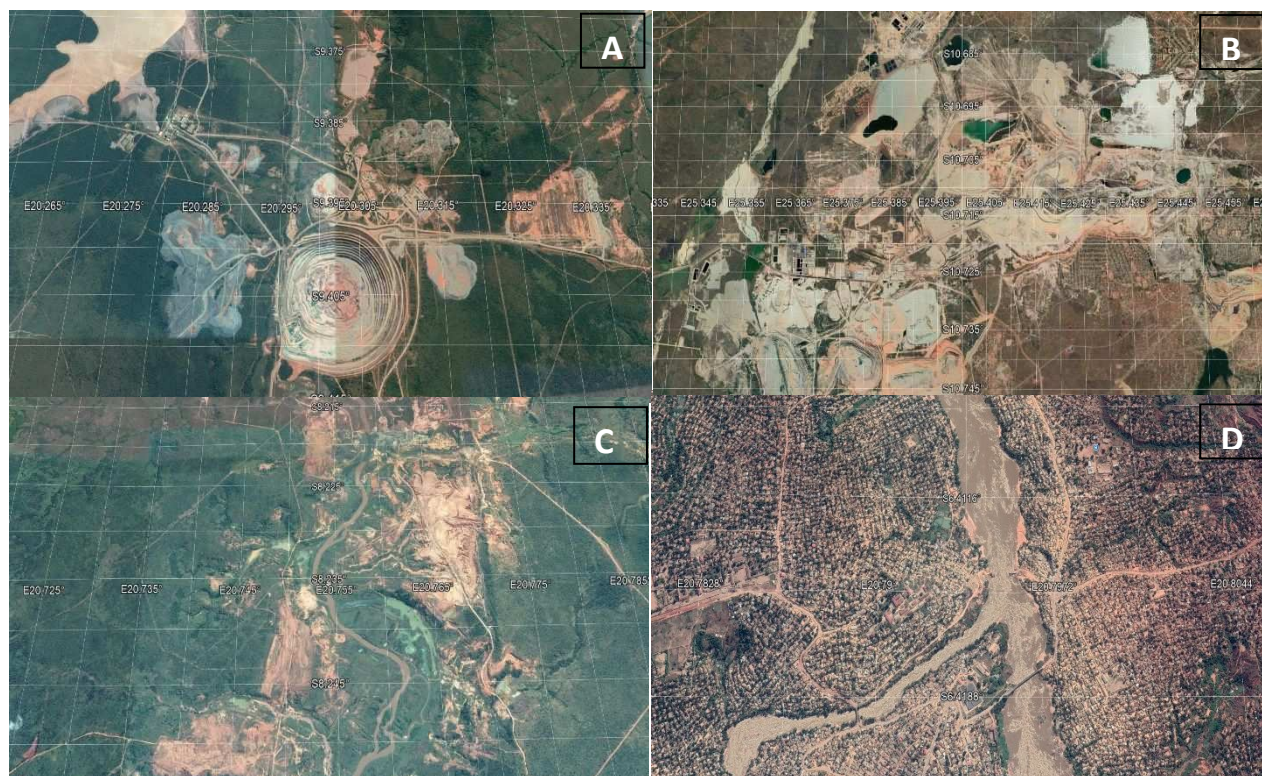


**Figure 4** Hierarchical levels of rivers in the CRB based on Strahler order



**Figure 5** An excerpt of the application of the approach for the Pool Malebo where the delineation points are positioned at the outlets and inlets of the contributing area. Blue dots represent the delineation points.





**Figure 6** Features of mining and urban areas included in catchments delineation (on top two industrial mining of Saurimo in Angola (A- catchment area: 7050 km<sup>2</sup>) and Kimoto in Katanga (B- catchment area: 5617 km<sup>2</sup>); bottom left an artisanal mining with dislocation of the river system in Camissombo, upper Kasai basin (C- catchment area: 14285 km<sup>2</sup>); and bottom right major town of Tshikapa in DRC (D- catchment area: 9189 km<sup>2</sup>)).

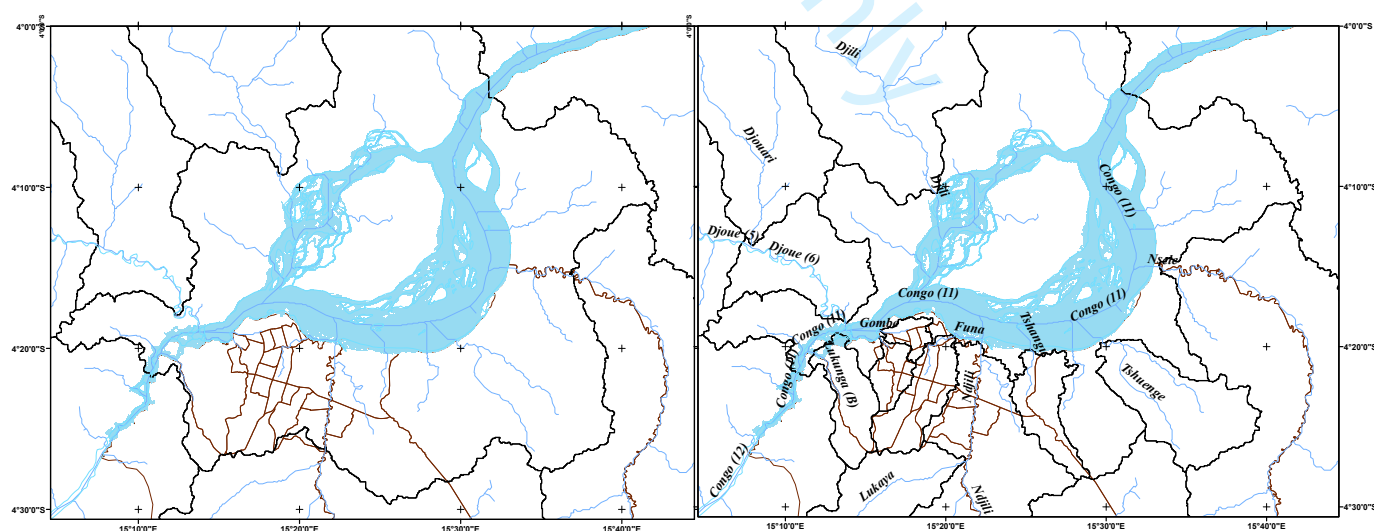


Figure 7 Illustration of second (Left) and third (Right) levels catchments delineation for the unit of Pool Malebo near the city of Kinshasa in DRC.

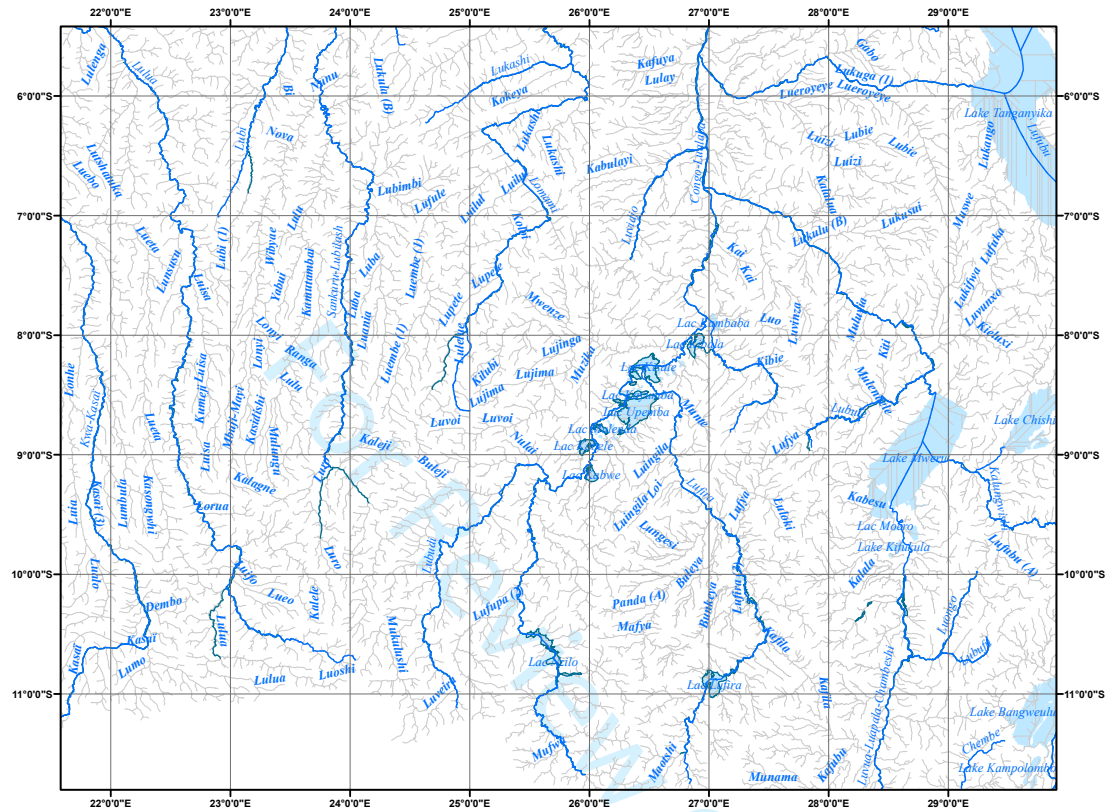


Figure 8 An excerpt of the identified river names within the CRB

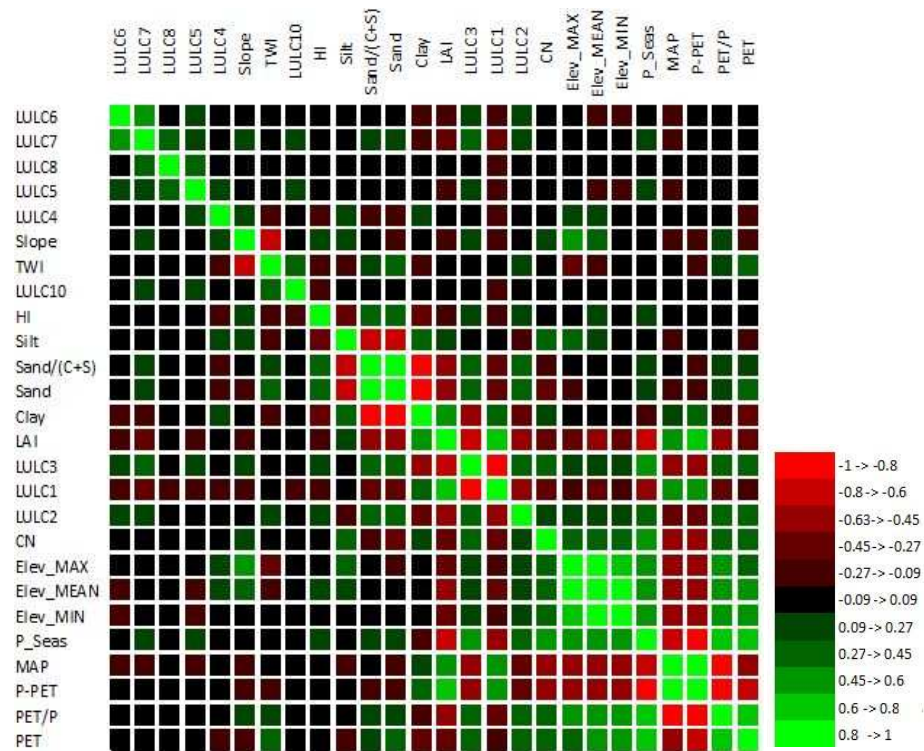


Figure 9 Pearson correlation matrix of the 26 variables of the physical basin properties. Refer to Table 1 for definition of the variables.

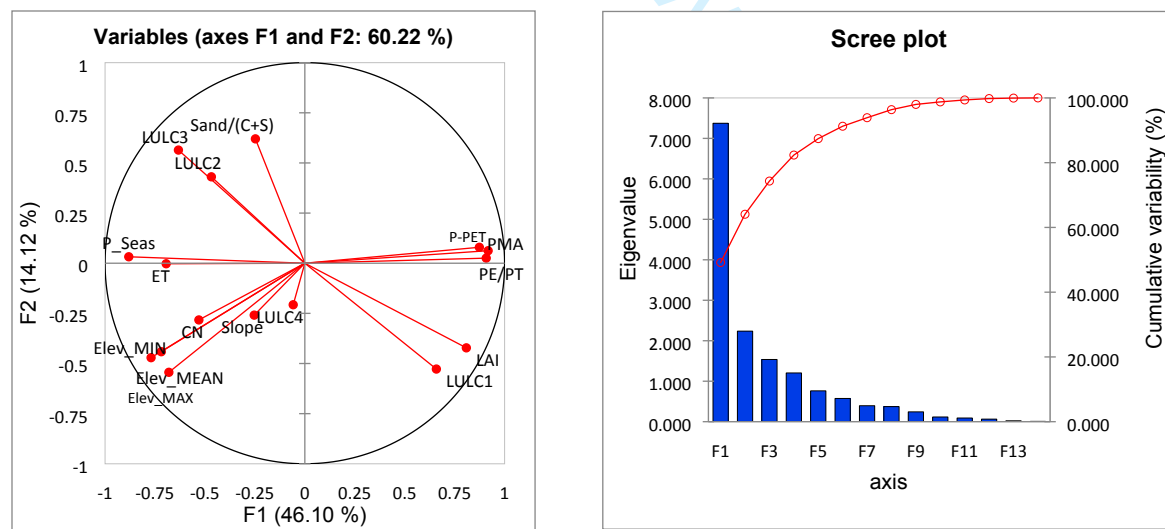


Figure 10 PCA ordination bi-plots for 15 variables physical basin attributes (left) and the scree plot of the Eigen value (right).



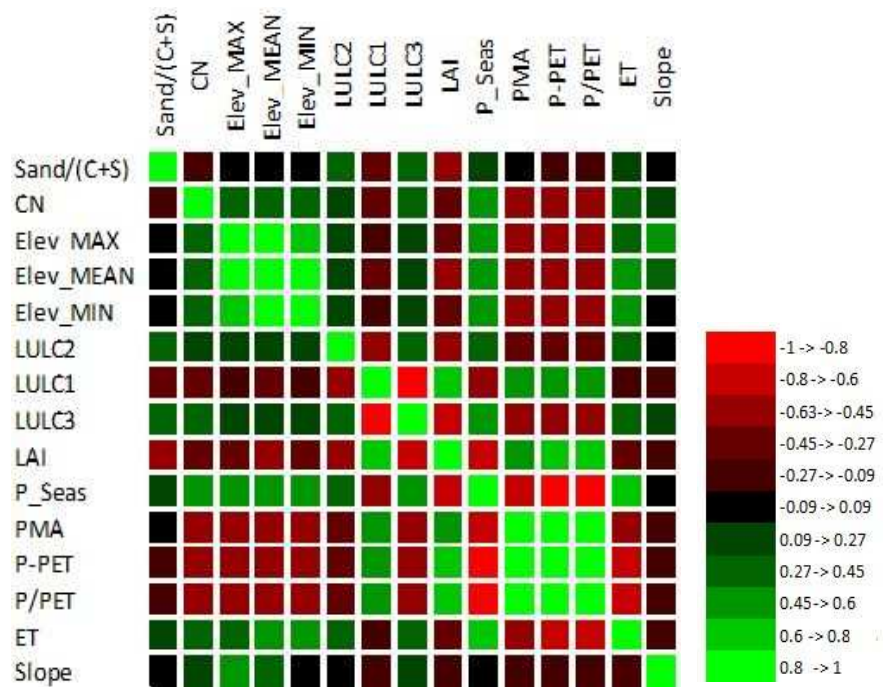


Figure 11 Correlation matrix (Pearson (n)) showing relationships between the 15 variables used in PCA.

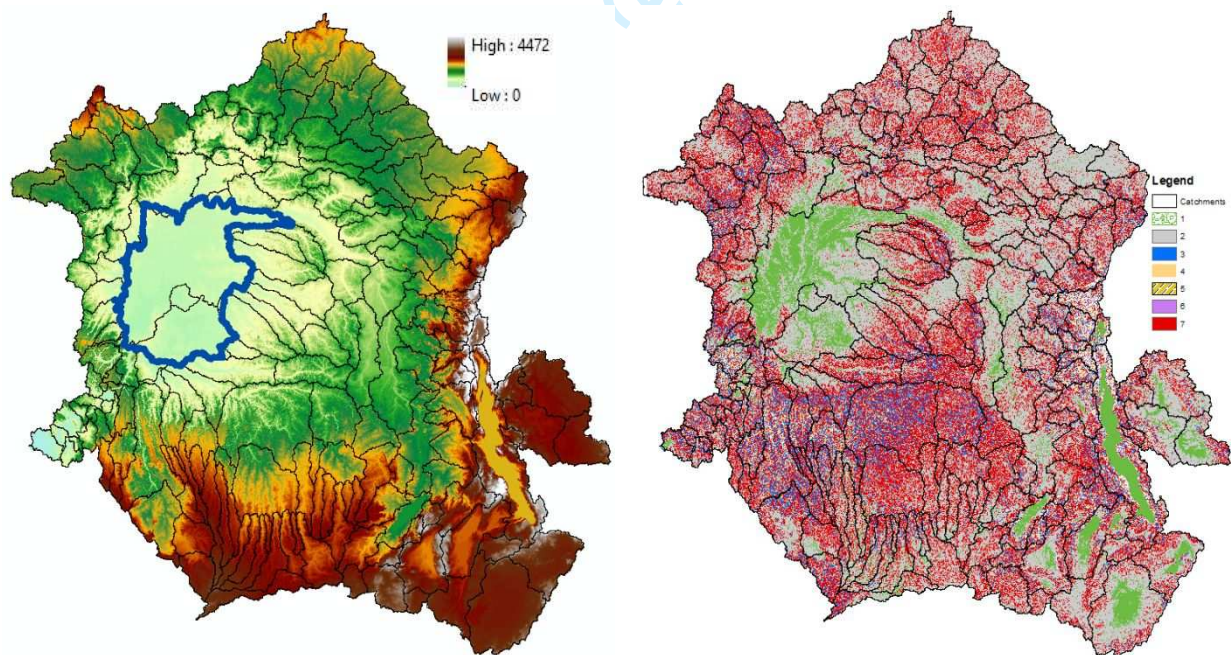


Figure 12 First level partition of catchment units along the elevation (left) and slope (right) gradients. The slope classes include 1 [0-0.25%], 2 [0.25-2%], 3 [2-5%], 4 [5-8%], 5 [8-15%], 6 [15-30%], 7 [30-84%].

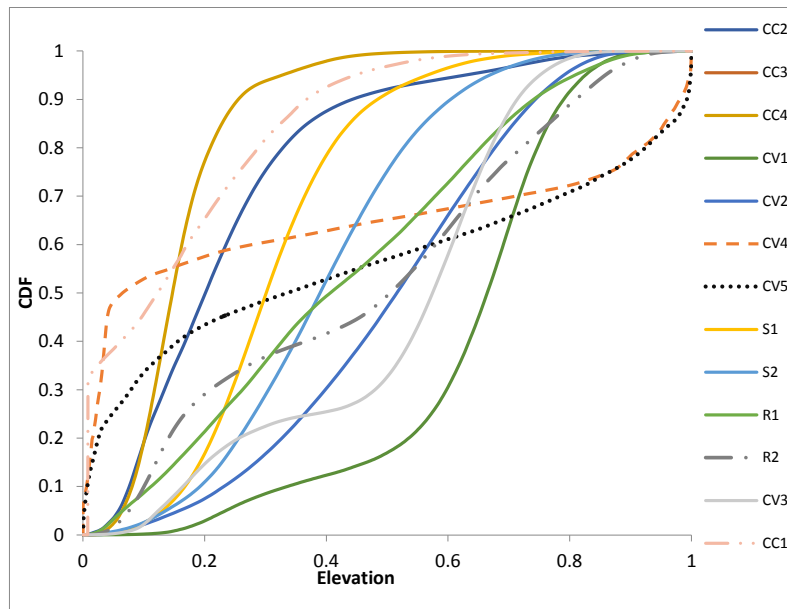


Figure 13 Cumulative Distribution Functions (CDF) forms across the dominant topographic gradients in the CRB [concave curves (CC), convex curves (CV), rectilinear curves (R), and S curves (S). Numbers refer to internal variant forms within the general shape].

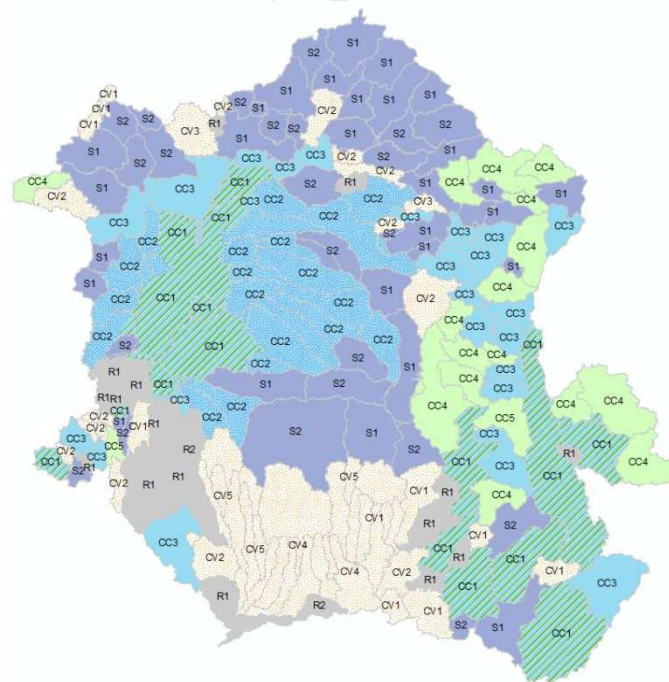


Figure 14 Patterns of different geomorphological groups and forms associated with the corresponding catchment units as derived from CDF. A single colour is used to represent sub-groups of the shapes S (S1, S2), R (R1, R2), CV (CV1,CV2,CV3,CV4, CV5)).



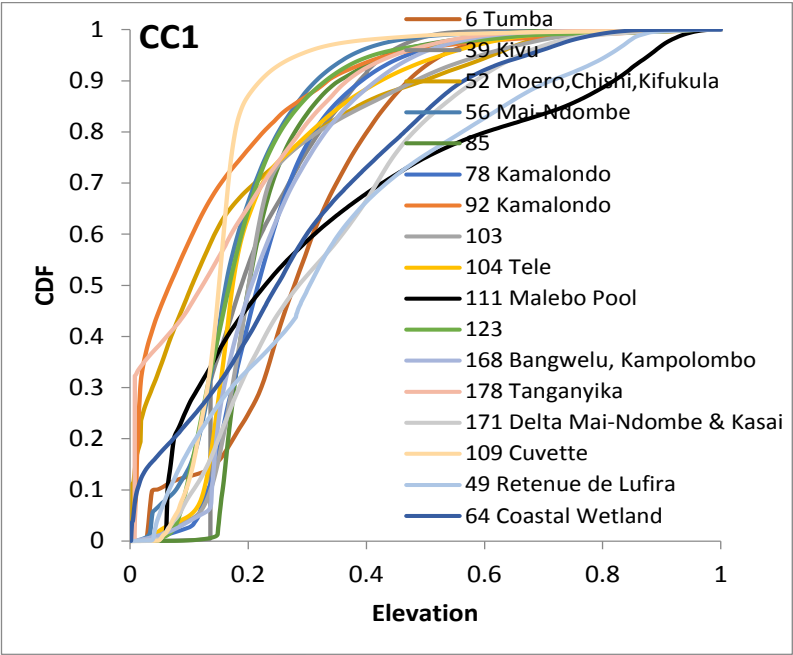


Figure 15 Variant forms of catchments characterized by concave shaped CDFs of the type of natural lakes and wetlands (refer to Figure 16 for the location of the feature ID)

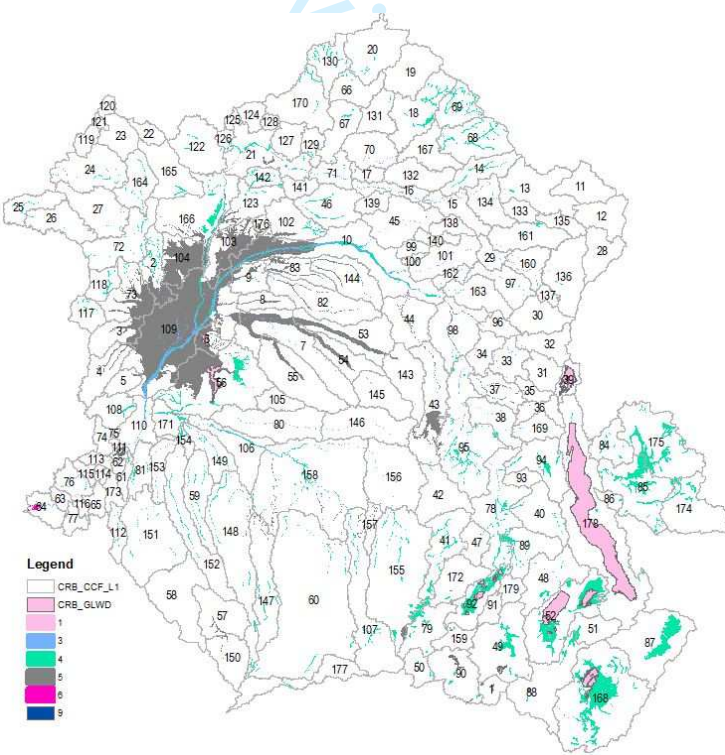


Figure 16 Map of the corresponding wetlands and natural lakes catchments [analysis done using 179 over 199 units]. 1: Lake, 3: Rivers, 4: Freshwater Marsh/Floodplain, 5: Flooded forest, 6: Coastal wetland, 9: Intermittent wetland/Lake.

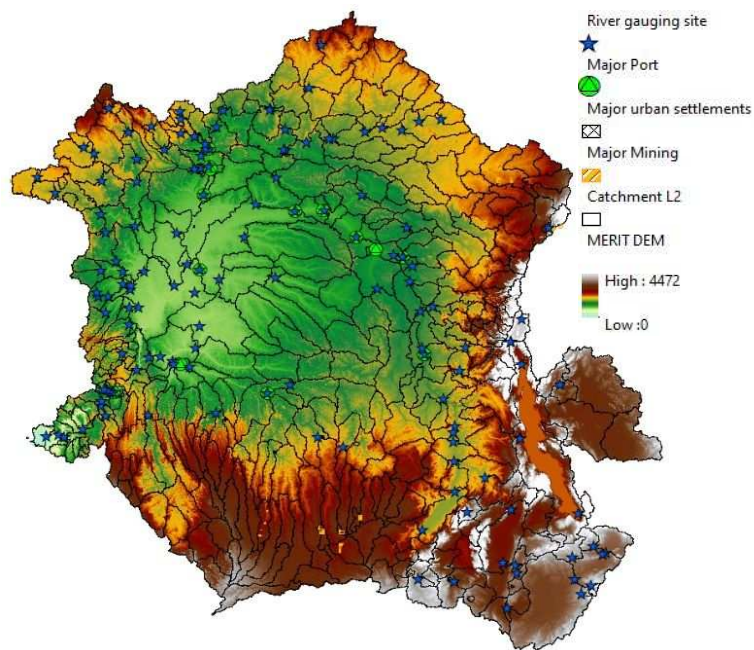


Figure 17 Spatial distribution of 403 catchment units including socio-economic and anthropogenic systems units

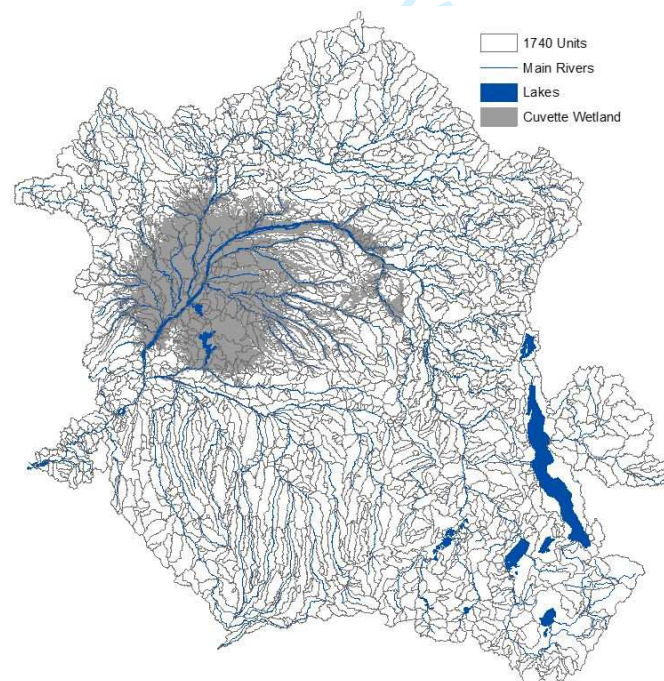


Figure 18 Spatial distribution of 1740 catchment units including the three-level partition

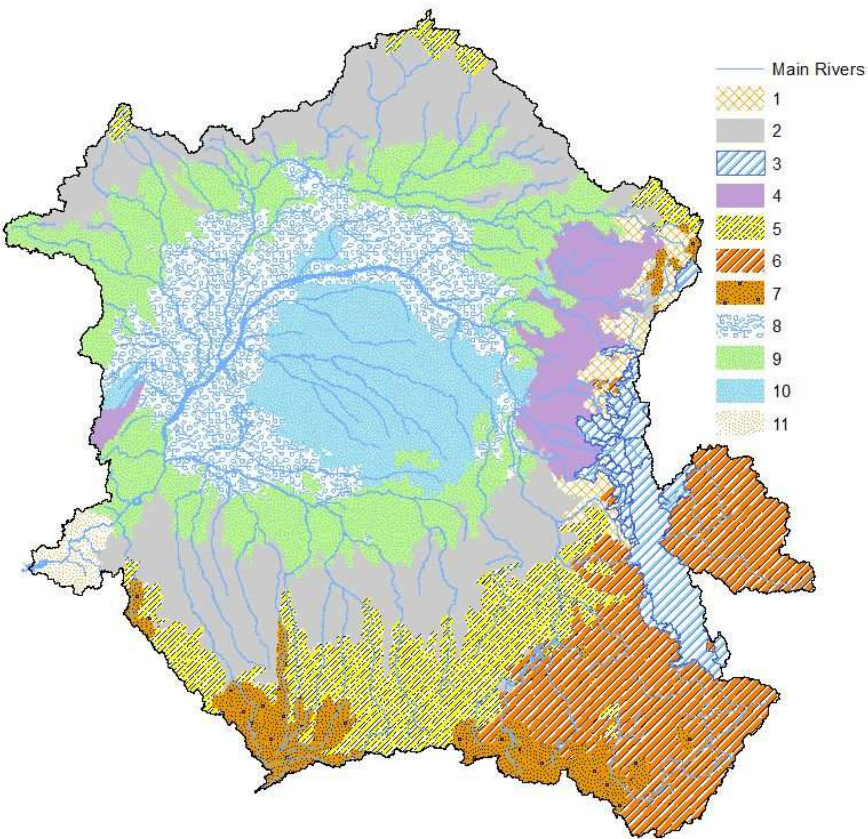


Figure 19      Graphical representation of 11 groups of homogenous physiographic setting as derived from the unsupervised classification based on HAC.

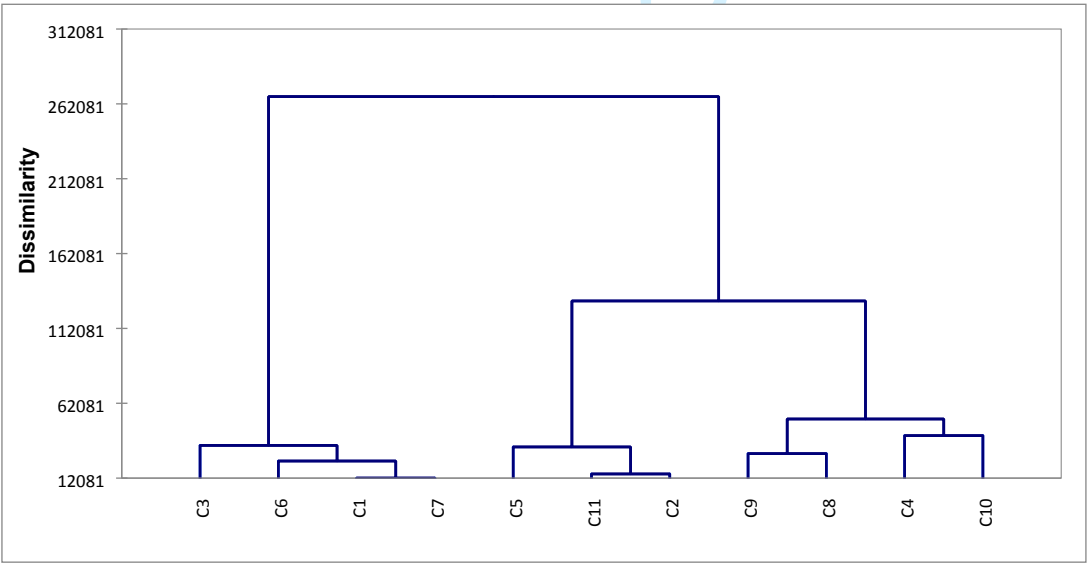


Figure 20      A dendrogram of the 11 groups of homogenous physiographic setting

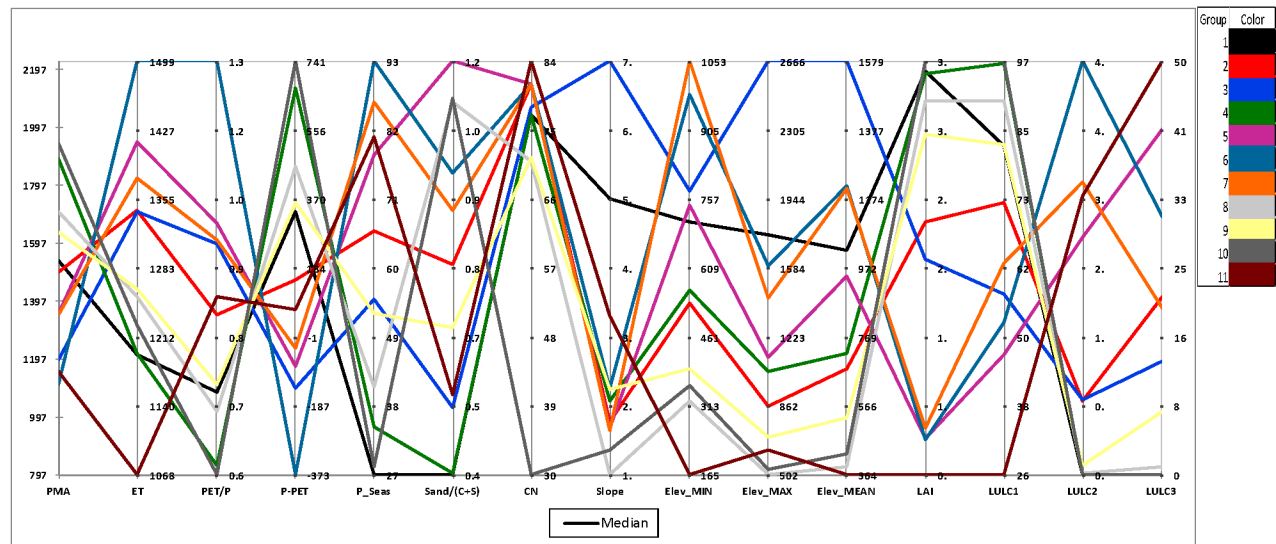


Figure 21 Parallel Coordinate Plot of the 11 groups of homogenous physiographic setting

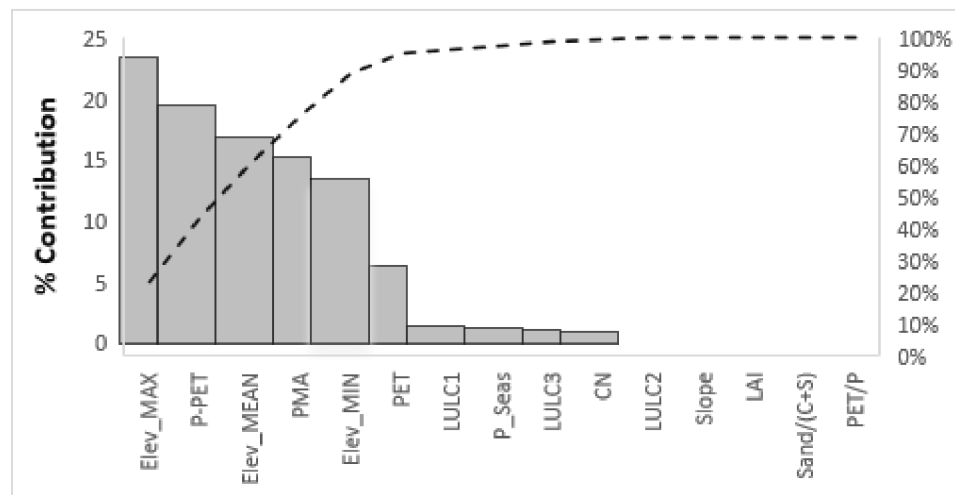


Figure 22 Overall contribution of the main variables to the average similarity within the regions of homogeneous physiographic settings.