2017

# The assessment of water resources in ungauged catchments in Rwanda

O. P. Abimbola
*University of Nebraska - Lincoln*

J. Wenninger
*Delft University of Technology*

R. Venneker
*UNESCO-IHE*

Aaron R. Mittelstet
*University of Nebraska-Lincoln*, amittelstet2@unl.edu

Journal of Hydrology: Regional Studies 13 (2017) 274–289

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

journal homepage: www.elsevier.com/locate/ejrh

# The assessment of water resources in ungauged catchments in Rwanda

O.P. Abimbola[a,c], J. Wenninger[a,b], R. Venneker[a], A.R. Mittelstet[c,*]

[a] UNESCO-IHE Institute for Water Education, Department of Water Science and Engineering, P.O. Box 3015, 2601 DA Delft, The Netherlands
[b] Delft University of Technology, Section of Water Resources, P.O. Box 5048, 2600 GA Delft, The Netherlands
[c] Biological Systems Engineering Department, University of Nebraska, 245 L.W. Chase Hall, Lincoln, NE 68586, USA

ARTICLE INFO

*Keywords:*
Ungauged catchments
Streamflow estimation
Stepwise multiple-regression
Catchment attributes

ABSTRACT

*Study region:* Rwanda is a landlocked country in Africa with precipitation ranging from 800 mm $yr^{-1}$ in the east to 1500 mm $yr^{-1}$ in high-altitude regions in the north and west.
*Study focus:* Streamflow estimation is an important task that is required in water resource assessments due to its importance in planning, decision-making and economic development. In this study, streamflow characteristics of ungauged catchments in Rwanda were calculated using a regionalization approach based on climate similarity and stepwise multiple-regression analysis. One climatic homogeneous region was identified and datasets of nine gauged stations and general available catchment characteristics were used to develop non-transformed and log-transformed regression models.
*New hydrological insights for the region:* Results of this study show that climate, physiography and land cover strongly influence the hydrology of catchments in Rwanda. Using leave-one-out cross-validation, the log-transformed models were found to predict the flow parameters more suitably. These models can be used for estimating the flow parameters in ungauged catchments in Rwanda and the methodology can be applied in any other region, as long as sufficient and good quality streamflow data is available.

## 1. Introduction

Assessment of water resources is of great value for national socio-economic development and stability of every country. Nevertheless, tools and data needed to carry out such assessments are often limited or lacking, especially in developing countries with limited technical capacity and funding (McNulty et al., 2016). At the core of the social and economic development of Rwanda is the aspiration of the country to become a middle-income country by the year 2020. This is formulated and described in *Rwanda Vision 2020* (MINECOFIN, 2000), and worked out into strategies, plans and actions for accelerated growth. This vision may not be a reality if a country-wide assessment of water resources is lacking.

Availability of water for food production is a major concern since Rwanda has an undulating topography throughout the country (RIWSP, 2012c). Thus, irrigation is practiced at small scales and is typically restricted to valley bottoms where 85–90% of the Rwandan population depends on subsistence farming. However, this may likely change in the future because a number of pilot studies in different parts of the country have been carried out to investigate the potential of rainwater harvesting for hill slope irrigation purposes (GOR, 2007).

In addition, shortages in power production due to a drop in water levels have caused interruption of the Rugezi hydropower plant.

* Corresponding author.
*E-mail address:* amittelstet2@unl.edu (A.R. Mittelstet).

http://dx.doi.org/10.1016/j.ejrh.2017.09.001
Received 9 May 2017; Received in revised form 8 August 2017; Accepted 1 September 2017
Available online 15 September 2017

Hence, as part of the national strategy to substantially increase power production, there is a need to increase the in-country hydropower production (GOR, 2007; MININFRA, 2011). Furthermore, the effects of water-related natural disasters, which are mostly connected with flood and drought conditions, cannot be over-emphasized. In general, more people are affected by droughts than flooding events in Rwanda (WFP, 2009).

In order to achieve *Rwanda Vision 2020*, there is need for a country-wide assessment so as to manage changes in water demand and supply (both surface and groundwater) for agricultural production, hydro-electric power generation, hazard mitigation and disaster preparedness.

Changing climatic conditions over longer periods of time affect water availability and have impact on the spatial and temporal variations of the water fluxes. An integrated water resources management relies on adequate water resources information that is acquired through continuous data collection, in combination with suitable analysis and assessment of the water-related information for water resources planning and development purposes (RIWSP, 2012c). In the case of Rwanda whereby such data or information is limited or lacking, the country would benefit significantly from the water resource assessment capabilities that hydrologic modelling can provide for predicting streamflow especially in ungauged catchments.

Predicting flow variables in ungauged or poorly-gauged catchments is one of the major concerns in hydrological studies, especially in regions with huge spatial variability of the hydrological environment and sparse or lack of data. In many parts of the world, current measurement networks are declining and the impacts of anthropogenic changes and climate amplify this issue. Hence, predictions of poorly gauged or ungauged catchments under these conditions are highly uncertain (Sivapalan et al., 2003).

Information from gauged catchments is usually transferred to the ungauged catchments using regionalization processes (Blöschl and Sivapalan, 1995). Studies on regionalization in hydrology have progressed continuously as a result of the need for streamflow predictions in ungauged catchments. Thus, understanding hydrological processes, their associated uncertainties and the development of models with increasing predictive power have become vital. In literature, a number of definitions of regionalization could be found, but the definition stated by Blöschl and Sivapalan (1995) is often used. They stated that "*regionalization is the process of transferring information from comparable catchments to the catchment of interest*". Since the objective of this study was to develop models to predict flow parameters in ungauged catchments in Rwanda, a regionalization approach was used. Some studies estimate parameters of streamflow statistics, usually flow quantiles, while others estimate parameters of rainfall-runoff models for simulating continuous streamflow or estimate continuous streamflow without using a model (Hrachowitz et al., 2013; He et al., 2011). There are many methods used for parameter regionalization (Merz and Blöschl, 2004).

The spatial proximity method assumes that catchment characteristics and climate vary smoothly in space. Thus spatial proximity, which is usually defined based on the distances between the catchment centroids or catchment outlets, between the catchments may be an appropriate measure of similarity when selecting the donor catchment (Li et al., 2009; Randrianasolo et al., 2011). A donor catchment is a catchment that is most similar in terms of its physiographic attributes to the catchment of interest (Parajka et al., 2005). In order to account for nested catchments, geostatistical distances can be used (Skoien et al., 2006; Skoien and Blöschl, 2007).

Similarity of catchment characteristics and climate, as an alternative method, selects the donor catchment(s) based on the similarity of the catchment characteristics and climate in the catchments. Similarity is calculated by the root mean square difference of all the characteristics in a pair of catchments (Blöschl et al., 2013). In order to make the characteristics comparable, they are usually standardized. Kokkonen et al. (2003) transferred the entire set of parameters from the catchment which has the most similar elevation to that of the catchment outlet while McIntyre et al. (2004) defined the most similar catchment on the basis of the catchment area, standardized annual mean precipitation and base-flow index. While some studies (Parajka et al., 2005; Zhang and Chiew, 2009) used a large number of catchment characteristics, others (e.g. Oudin et al., 2010) used fewer, yet more relevant catchment characteristics.

The model averaging method uses a weighted combination of the parameter sets from more than one donor catchment, where the catchments are chosen either based on spatial proximity, catchment characteristics or both (Seibert and Beven, 2009). Each catchment can either be assigned to its own group of donor catchments or, alternatively, the region can be divided into groups of catchments (Burn and Boorman, 1993).

Parameter regression is the most widely used method for rainfall-runoff model regionalization (McIntyre et al., 2004). This method relates the model parameters explanatorily to physiographic characteristics in the gauged catchments through empirical equations which can then be used to predict the model parameters in the ungauged catchments (Merz and Blöschl, 2004; Mazvimavi et al., 2004, 2005; Wagener and Wheater, 2006; Young, 2006; Parajka et al., 2013). To investigate the value of seasonality indices for regionalizing low flows, Laaha and Blöschl (2006, 2007) used stepwise-multiple regressions based on physical catchment characteristics and seasonality indices to make regionalization models. Using cross-validation, they assessed the value of different models that incorporate seasonality by different approaches in order to predict low flows in ungauged catchments. They compared the models for the 95% quantile of specific discharges and also examined the specific low flow discharges of the summer and winter periods ($q_{95s}$, $q_{95w}$). Their results showed that grouping the study area into different regions and separate regressions in each region provides the best model performance. According to Laaha and Blöschl (2006, 2007), a global regression model yields the lowest performance and a global regression model that uses regional calibration coefficients only performs slightly better. They recommended that separate regression models in each of the regions are to be chosen over a global model in order to represent differences in the way catchment characteristics are related to low flows.

In order to make reliable predictions in ungauged basins, it is preferable that the equations which relate the model parameters and the catchment characteristics should be hydrologically reasonable. According to Sefton and Howarth (1998), this is not always possible because the explanation of the regression equations is often not straightforward by reason of unrepresentative catchment characteristics and issues related to the selection of model parameters (Blöschl, 2005). As pointed out by Kokkonen et al. (2003), high significance of regression models does not necessarily give a set of parameters with a good predictive power. Hence, interpreting the physical meaning of regression relationships amongst model parameters and characteristics needs careful consideration.

Model averaging and parameter regression can also be applied simultaneously by calibrating the coefficients of these relationships as an alternative to first estimating model parameters at each gauged catchment and then relating them to the catchment characteristics by an empirical equation. This permits the finding of more reliable parameters compared to merely calibrating the model parameters themselves and drawing on the spatial data contained within the catchment characteristics (Parajka et al., 2013). Rather than using only one method, some studies compare regionalization methods for estimating the model parameters in ungauged catchments (e.g., Merz and Blöschl, 2004).

The objective of this study is to assess the spatial characteristics of water resources of gauged catchments in Rwanda and to regionalize the information to estimate, in an optimal way, flow quantiles (mean, low and high flows) and the slope of flow duration curve in ungauged catchments. The aim is to gain more knowledge from the differences and similarities between catchments, and to interpret the differences in terms of the underlying climate and landscape characteristics. This study addresses (1) the extent to which prediction performance correlate with climate and catchment characteristics, (2) which type of stepwise-parameter regression performs better (non-transformed or log-transformed regression), (3) how hydrologically meaningful the explanatory variables chosen by the forward stepwise regression procedure are, and (4) how model complexity impacts performance. The regression models developed will contribute more towards understanding the water fluxes of the catchments, and also lead to a better design of integrated water resources management plans.

## 2. Material and methods

### 2.1. Study site and datasets

This study was carried out in Rwanda using data from from 1961 and 2013. Rwanda is a land-locked country with a surface area of 26,338 km$^2$, of which 2165 km$^2$ is water. It is a mountainous country with over 70% of the land surface having slopes greater than 10%. Elevation ranges from 950 m above sea level (m ASL) at Rusizi River in the southwest to 4507 m ASL at Volcan Karisimbi in the northwest (Fig. 1). The topography is complex and composed of interlocking rolling hills in most parts of the country (McSweeney, 2011; Museruka et al., 2011). The Congo-Nile Ridge, which is a range of mountains with an altitude ranging between 2500 and 4500 m ASL, forms the drainage divide between the Congo River basin, occupying 20% of the area in the West, and the Nile River basin covering 80% of the area in the East (RIWSP, 2012c).

Despite being located in the tropical belt, Rwanda experiences a temperate climate due to its high elevation. Mean precipitation ranges from 800 mm yr$^{-1}$ in the east to 1500 mm yr$^{-1}$ in high-altitude regions in the north and west (RIWSP, 2012c). The lowest mean temperatures, 15 °C, are observed in the high altitude ranges of the Nile-Congo drainage divide and in the northern mountainous regions. Moderate mean temperatures ranging between 18 and 20 °C are found in the central plateau of the country. The highest mean temperatures between 20 and 24 °C are recorded in the lowland undulating plains in the eastern and south-central regions (McSweeney, 2011; Museruka et al., 2011; RIWSP, 2012c). Mean annual potential evapotranspiration ranges from less than 1000 mm in the northwest to more than 1400 mm in the southern border margin (RIWSP, 2012c) as shown in Fig. 2. Rwanda has quite a number of meteorological stations but with relatively fewer stream gauging stations. About half of the stream gauging stations have only stage measurements with sparse or no discharge measurements. The spatial distributions of the hydrological and meteorological stations as well as the boundaries of nine stream gauged catchments with sufficient and good quality streamflow data are shown in Fig. 3. For this study, the six major land cover classifications used were: cropland, grassland, shrubland, forest, salt hardpan and water bodies (RIWSP, 2012c). Climatological stations are those that measure and record rainfall as well as surface air temperature and some other parameters. Precipitation stations are those which measure and record daily rainfall only. Agro-synoptic and synoptic stations are those which observe and record all the surface meteorological data such as precipitation, minimum and maximum temperature, wind speed and direction, relative humidity, solar radiation, clouds, atmospheric pressure, sunshine hours, evaporation and visibility. Agro-synoptic stations are run by Rwanda Meteorology Agency.

Stage and/or discharge records of only 17 stations are documented by RWRIS in HydroScape Version 2.0 (RIWSP, 2012b) as shown in Table 1. These records were evaluated, cleaned from obviously erroneous data and, where possible, corrected by checking for reading or typing errors of individual observations, manually filled parts in the record, consistency in seasonal variability, abrupt shifts in stage, and gauge datum shifts as well as range changes after record gaps. In addition, the discharge rating curves were checked. For some of the stations, it was necessary for the rating curves and the existing discharge records to be discarded while for the other stations they were either acceptable or revised from the discharge measurements (RIWSP, 2012a). This resulted in nine stations with 4–51 years of quality record between February 1961 and January 2013 and an average record period of 35 years. Flow data from neighboring countries were not available as at the time of this study hence, analysis was limited to flow records from Rwanda only.

Two sets of catchments were used in this study. The first was a set of nine gauged catchments which was used in estimating selected parameters and developing models for prediction. The second set included 68 ungauged catchments (based on Level 2 of the Pfafstetter Coding system) to which the prediction models were applied. The physiographic and hydro-climatic datasets for the nine gauged catchments (referred to by the names of the stream gauging stations) and 68 ungauged catchments were obtained from RWRIS HydroScape 2.0 as shown in Fig. 4 (RIWSP, 2012b). The RWRIS HydroScape was also used to define several catchments characteristics (Table 2). Variability of catchment characteristics for all gauged and ungauged catchments is shown in Table 3.

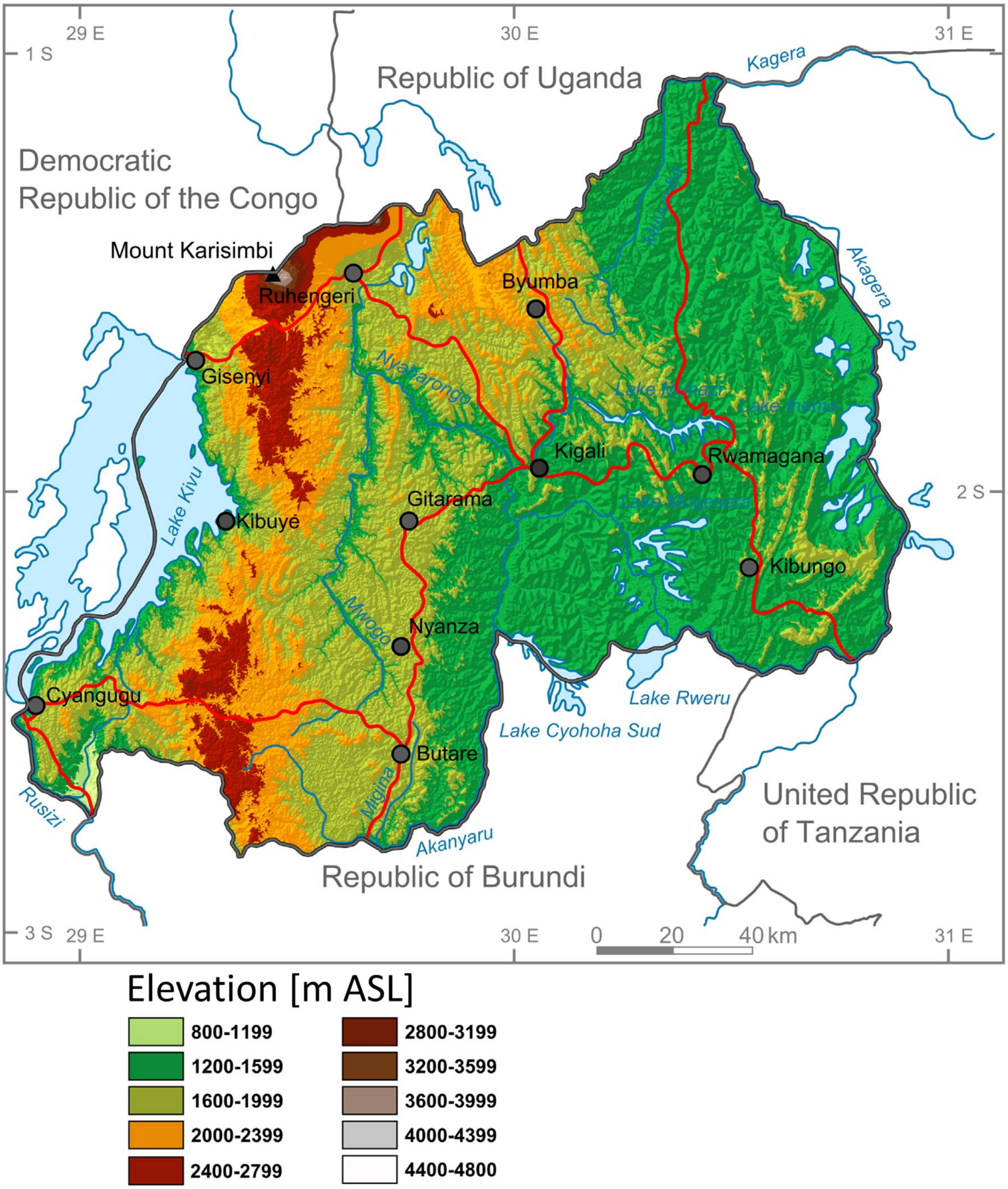


**Fig. 1.** Rwanda DEM (RIWSP HydroScape, 2012b) (m ASL is meters above sea level).

### 2.2. Methodology

In order to compute the mean areal water balance distribution, an estimate of the spatially distributed rainfall was carried out to obtain mean annual rainfall for the entire domain covered by the sub-basins shown in Fig. 4. Annual rainfall for all gauged and ungauged sub-basins was calculated using GIS tools and FAO Local Climate Estimator (New_LocClim) which was developed to provide an estimate of climatic conditions at locations for which no observations are available. The values were compared with the corresponding estimates for the rainfall stations, and showed that the estimated values agree within 10% or better. Potential evapotranspiration was also calculated for all sub-basins using GIS tools and FAO New_LocClim. Actual evapotranspiration was directly obtained as the difference between precipitation and runoff based on the principle that the change in storage over long periods of

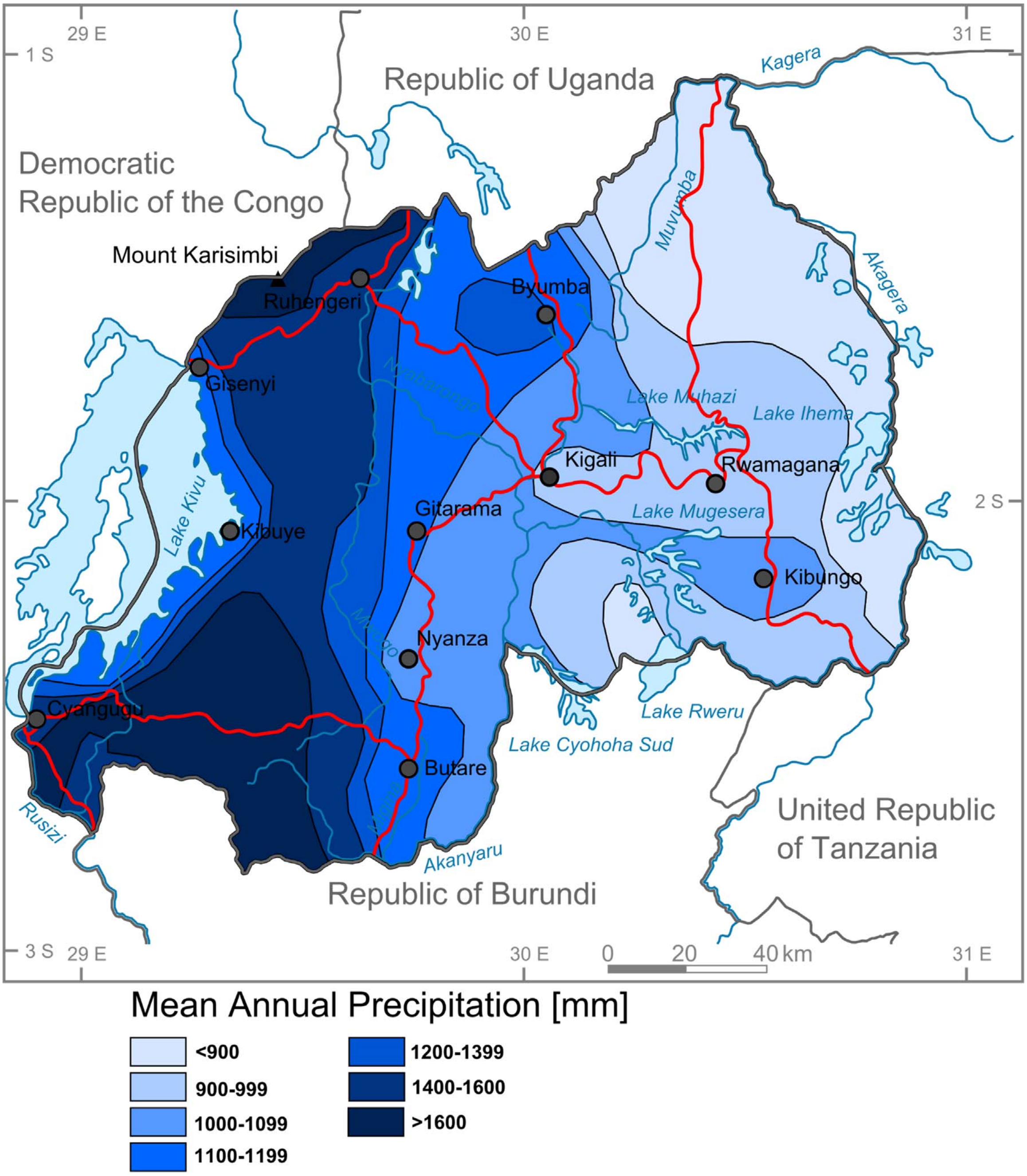


**Fig. 2.** Spatial distribution of mean annual precipitation (RIWSP, 2012c).

time equals zero (RIWSP, 2012d).

Climatic classification was conducted for the gauged catchments based on the hydro-climatic region in which catchments were located using a Budyko curve approach. Budyko (1974) used the long-term average water and energy balance variables to develop a climatic classification scheme. According to Gerrits et al. (2009), the Budyko curve is often used in water resources studies to predict evaporation as a function of dryness index ($ET_{pot}/PREC$). Fig. 5 shows the application of the Budyko climatic classification method to compare the nine gauged catchments. The climate of each of these catchments is presented on the Budyko curve, which is a plot of $ET_{act}/PREC$, the ratio of average annual actual evapotranspiration ($ET_{act}$) to average annual precipitation ($PREC$) as a function of the dryness index. Actual evapotranspiration ($ET_{act}$) for each catchment was calculated as the long-term difference between precipitation and runoff for the nine catchments. This assumes all recharge discharged to the stream. The lines on the Budyko curve represent globally averaged results obtained by Schreiber (1904), Pike (1964) and Budyko (1974). Three climatic thresholds were used for Budyko analysis of the gauged catchments. A dryness index between 0 and 0.7 was considered wet while a dryness index between 1.3

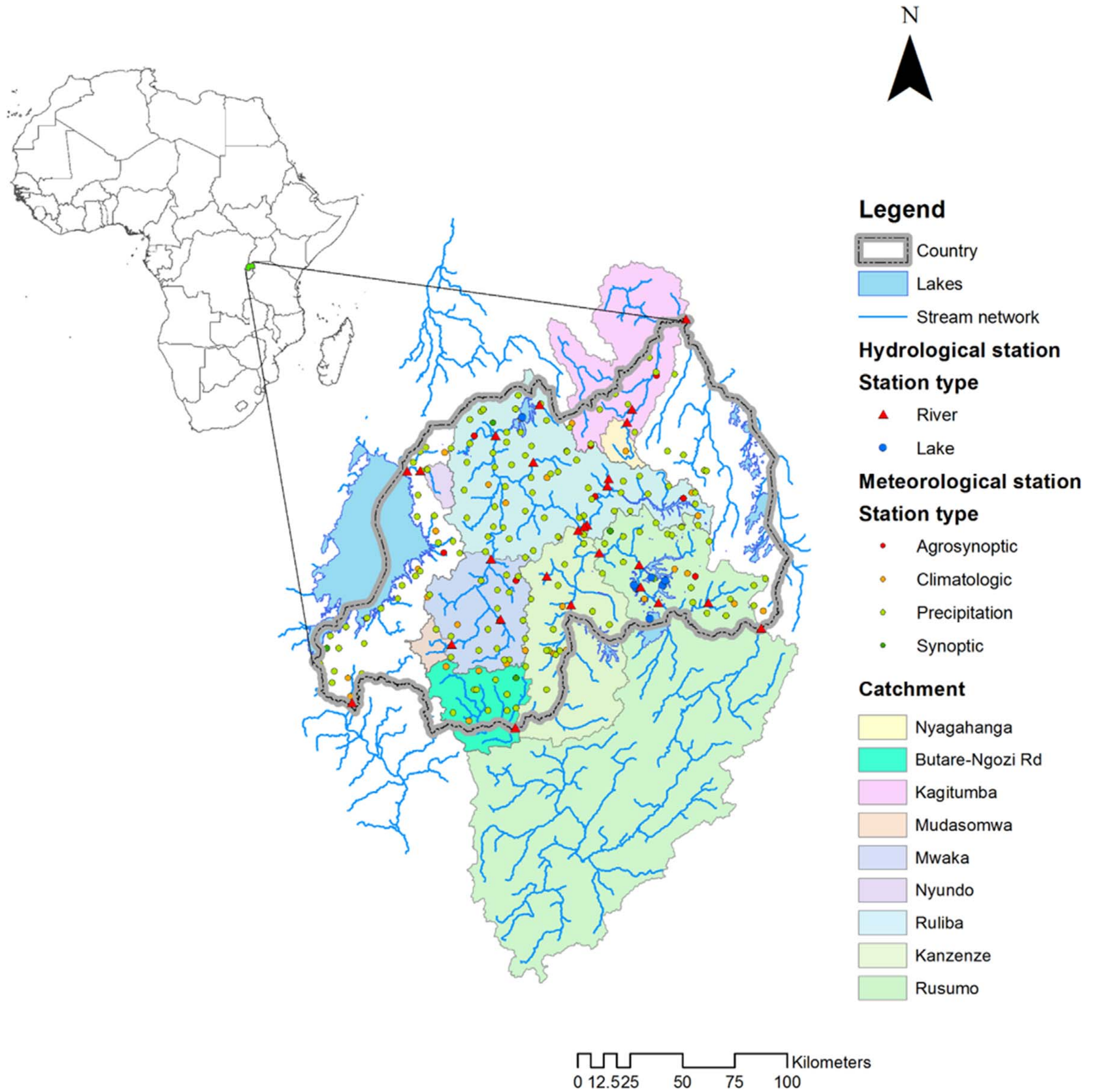


**Fig. 3.** Locations of the gauging and climate stations in Rwanda with gauged catchments boundaries.

and infinity was considered dry. Values between 0.7 and 1.3 were considered medium category (i.e. not too wet nor too dry). It can be seen from the figure that all the catchments are in the medium category with dry index values between 0.86 and 1.18. Although all nine gauged catchments fall within the medium category, they were only used in one pool for regionalization analysis because they have comparable evaporative indices (between 0.56 and 0.88). This implies similarity of climate in the catchments with respect to relative water and energy availability and hence no need for using different regional equations to predict model parameters for wet and dry catchments. Thus, all nine gauged catchments were selected as donor catchments and the entire parameter set was transferred from these catchments to the ungauged catchments.

In order to predict flow magnitudes in ungauged catchments of Rwanda, the mean-daily, 5-, 50-, and 95-percent-flow duration ($Q_{mean}$, $Q_5$, $Q_{50}$, $Q_{95}$,) in addition to the slope of the flow duration curve ($FDC_{slp}$) were estimated for the nine gauging stations. To make the flow quantile values more comparable across scales, specific high flow discharges $q_i$ (l/s·km$^2$) were computed by standardizing $Q_i$ values by respective catchment areas (Table 4). The shape of a FDC gives a measure of variability and is determined by hydrologic and geologic characteristics of the catchment. Fig. 6 shows the flow duration curves (plotted on a log-normal scale) for the gauged catchments. Slopes of the flow duration curves ($FDC_{slp}$) were estimated from the standard deviation of the logarithms of the discharges.

After determining the flow parameters (flow quantiles and index) and physiographic characteristics of the selected gauged catchments, a multiple regression analysis, which is a statistical approach for investigating the relationship between a dependent and

**Table 1**
Summary of discharge data for all stream gauging stations .

| Station name | Data acquisition | Data update | Record start | Record end | Hydrometric observation | | Percent missing | Record period (years) |
|---|---|---|---|---|---|---|---|---|
| | | | | | Flow | Stage | | |
| Bugarama | 07/01/2013 | 09/01/2013 | 03/01/1974 | 21/12/2012 | | Y | 61% | 38 |
| Nyundo | 05/01/2013 | 09/01/2013 | 02/01/1974 | 31/12/2012 | Y | Y | 54% | 38 |
| Kagitumba | 05/01/2013 | 09/01/2013 | 03/01/1974 | 01/01/2013 | Y | Y | 61% | 38 |
| Nyagahanga | 07/01/2013 | 09/01/2013 | 02/01/1983 | 31/12/2012 | Y | Y | 61% | 29 |
| Rusumo | 15/12/2012 | 21/12/2012 | 02/01/1970 | 24/08/2012 | Y | Y | 44% | 42 |
| Mfune | 07/01/2013 | 09/01/2013 | 02/01/1971 | 31/12/2012 | | Y | 26% | 41 |
| Rwinzoka | 07/01/2013 | 09/01/2013 | 09/09/2008 | 01/01/2013 | | Y | 3% | 4 |
| Kanzenze | 05/01/2013 | 09/01/2013 | 08/03/1971 | 30/12/2012 | Y | Y | 57% | 41 |
| Gihinga | 07/01/2013 | 09/01/2013 | 02/01/1974 | 31/12/2012 | | Y | 59% | 38 |
| Butare-Ngozi Rd | 05/01/2013 | 09/01/2013 | 02/01/1971 | 31/12/2012 | Y | Y | 37% | 41 |
| Ruliba | 05/01/2013 | 09/01/2013 | 02/01/1961 | 31/12/2012 | Y | Y | 32% | 51 |
| Nemba | 07/01/2013 | 09/01/2013 | 17/05/1972 | 31/12/2012 | | Y | 60% | 40 |
| Nyakinama | 07/01/2013 | 09/01/2013 | 08/07/1995 | 01/01/2013 | | Y | 42% | 17 |
| Mwaka | 05/01/2013 | 09/01/2013 | 02/10/1971 | 31/12/2012 | Y | Y | 42% | 41 |
| Mudasomwa | 07/01/2013 | 09/01/2013 | 25/08/1987 | 31/12/2012 | Y | Y | 60% | 24 |
| Nyabisindu | 07/01/2013 | 09/01/2013 | 07/03/1972 | 31/12/2012 | | Y | 47% | 40 |
| All stations | | | 02/01/1961 | 01/01/2013 | | | | |

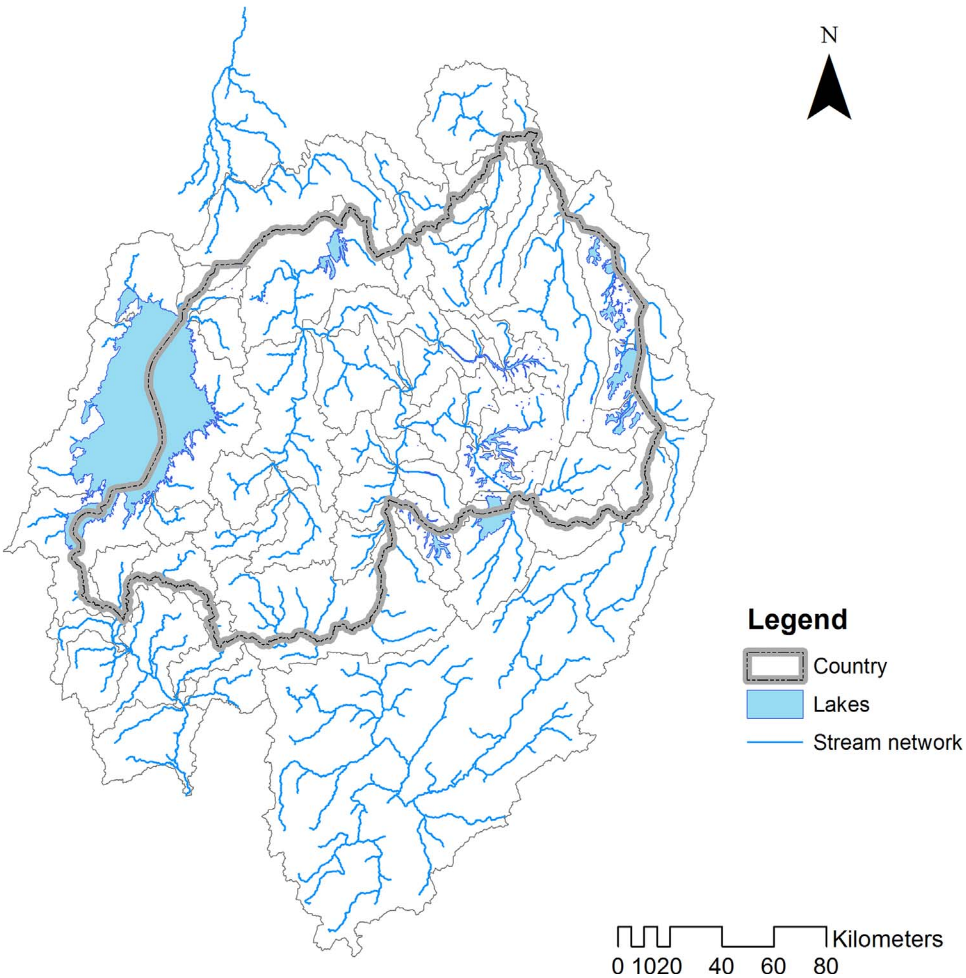


**Fig. 4.** Ungauged catchments in Rwanda.

**Table 2**
Abbreviations and units of all catchment characteristics for multiple regressions .

| Symbol | Characteristic description | Unit |
|---|---|---|
| Physiographic | | |
| $A$ | Catchment area | $km^2$ |
| $C_P$ | Catchment perimeter | km |
| $H_m$ | Mean catchment elevation | m |
| $H_+$ | Maximum catchment elevation | m |
| $H-$ | Minimum catchment elevation | m |
| $H_R$ | Range of catchment elevation | m |
| $C_S$ | Mean catchment slope | % |
| $D$ | River density | $km/km^2$ |
| Climatic | | |
| $PREC$ | Mean annual precipitation | mm |
| $ET_{pot}$ | Mean annual potential evapotranspiration | mm |
| Hydrologic | | |
| $q_p$ | Streamflow ($p$ = 5%, 50%, 95% and mean) | $l/s\ km^2$ |
| $FDC_{slp}$ | Slope of flow duration curve | (−) |
| Land cover | | |
| $L_C$ | Percent of cropland | % |
| $L_G$ | Percent of grassland | % |
| $L_S$ | Percent of shrubland | % |
| $L_F$ | Percent of forest | % |
| $L_{HP}$ | Percent of salt hardpan | % |
| $L_W$ | Percent of water bodies | % |

**Table 3**
Summary of catchment characteristics for gauged and ungauged catchments .

| | Gauged | | | Ungauged | | |
|---|---|---|---|---|---|---|
| | Min | Mean | Max | Min | Mean | Max |
| Area ($km^2$) | 215 | 6811 | 30644 | 10 | 780 | 12232 |
| Perimeter (km) | 75.00 | 436.00 | 1120.00 | 14 | 141 | 735 |
| Mean Elevation (m) | 1642 | 1934 | 2419 | 979 | 1620 | 2295 |
| Minimum Elevation (m) | 1284 | 1487 | 2023 | 771 | 1280 | 1792 |
| Maximum Elevation (m) | 2225 | 3295 | 4468 | 1491 | 2356 | 4477 |
| Range Elevation (m) | 829 | 1809 | 3127 | 238 | 1076 | 3334 |
| Slope (%) | 16 | 20 | 28 | 5 | 16 | 38 |
| Precipitation (mm) | 996 | 1196 | 1320 | 907 | 1117 | 1511 |
| ETpot (mm) | 1137 | 1235 | 1327 | 1128 | 1257 | 1434 |
| River Density ($km/km^2$) | 0.22 | 0.25 | 0.26 | 0.17 | 0.27 | 0.38 |
| Forest (%) | 0.20 | 25.60 | 81.70 | 0.00 | 22.90 | 100.00 |
| Shrubland (%) | 0.00 | 19.00 | 39.40 | 0.00 | 3.45 | 34.75 |
| Grassland (%) | 0.00 | 3.50 | 12.70 | 0.00 | 71.12 | 100.00 |
| Cropland (%) | 5.50 | 51.50 | 99.80 | 0.00 | 1.13 | 23.94 |
| Salthardpans (%) | 0.00 | 0.00 | 0.20 | 0.00 | 0.04 | 1.21 |
| Waterbodies (%) | 0.00 | 0.40 | 1.40 | 0.00 | 0.01 | 0.83 |

multiple independent (explanatory) variables, was used. The most widely used multiple regression equations are based on linear relationships although different relationships such as a logarithmic equation are also used. Eqs. (1) and (2) show examples of linear non-transformed and linear log-transformed multiple regressions based on three independent variables. In this study both non-transformed and log-transformed, which avoids heteroscedasticity and non-normality of the residuals of the regressions, were used.

$$Y' = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \quad (1)$$

$$\log(Y') = \log(\beta_0) + \beta_1 \log(X_1) + \beta_2 \log(X_2) + \beta_3 \log(X_3) \quad (2)$$

Eq. (2) can be re-written as

$$Y' = 10^{\beta_0} \times X_1^{\beta_1} \times X_2^{\beta_2} \times X_3^{\beta_3} \quad (3)$$

Where $Y'$ is the estimated dependent variable by the regression equation, $\beta_0$ is the intercept which is a constant value, and $\beta_i$ ($i$ = 1, 2, 3) are the regression coefficients which assign the effects of the independent variables $X_i$ on the dependent variable.

Regression models were fit to all gauged catchments for predicting the flow quantiles ($q_5$, $q_{50}$, $q_{mean}$, $q_{95}$) and index ($FDC_{slp}$). This was done based on the catchment characteristics, and in ascending order of the number of explanatory variables added by forward

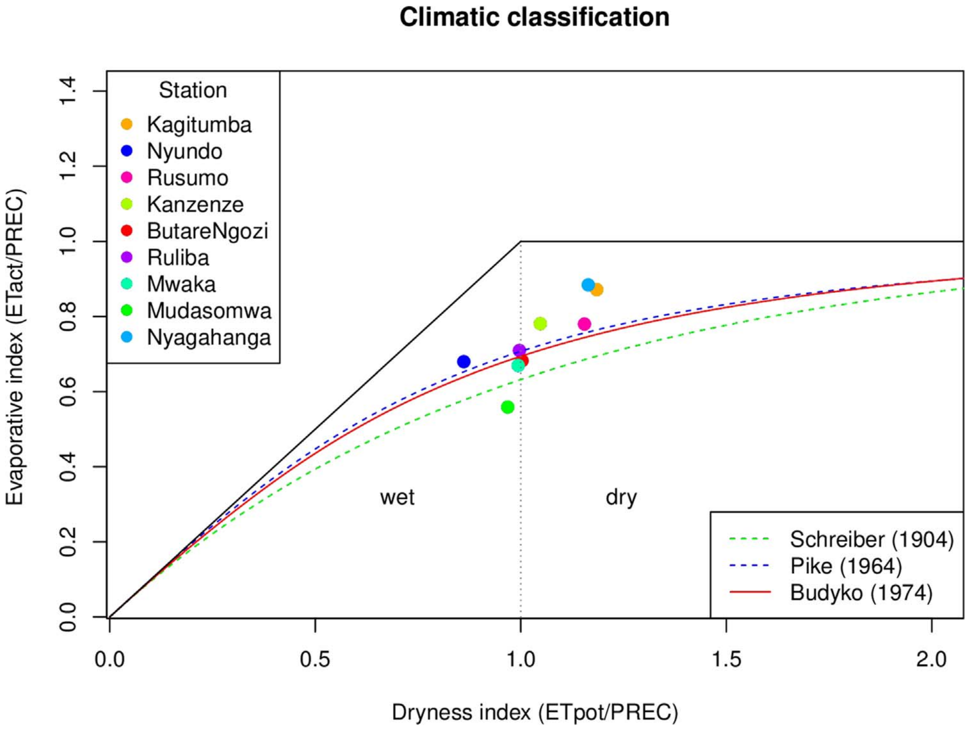


**Fig. 5.** Climatic classification based on the observations from nine gauged catchments. The 1:1 line defines the available energy limit to evapotranspiration ($ET_{pot} < PREC$), while the horizontal line defines the available water limit ($ET_{act} < PREC$).

**Table 4**
Flow quantiles and index of gauged catchments.

| Location | $q_5$ (l/s km$^2$) | $q_{50}$ (l/s km$^2$) | $q_{mean}$ (l/s km$^2$) | $q_{95}$ (l/s km$^2$) | $FDC_{slp}$ (−) |
|---|---|---|---|---|---|
| Kagitumba | 9.2 | 3.3 | 4.04 | 1.71 | 0.22 |
| Nyundo | 24.2 | 11.6 | 13.39 | 8.65 | 0.14 |
| Rusumo | 13.8 | 7.4 | 8.01 | 4.95 | 0.13 |
| Kanzenze | 14.2 | 7.6 | 8.34 | 4.91 | 0.14 |
| Butare-Ngozi Rd | 27.6 | 11.1 | 13.22 | 5.17 | 0.23 |
| Ruliba | 20.8 | 9.9 | 11.13 | 6.01 | 0.16 |
| Mwaka | 26.3 | 11.1 | 13.09 | 6.87 | 0.18 |
| Mudasomwa | 41.0 | 14.2 | 18.06 | 8.04 | 0.21 |
| Nyagahanga | 11.8 | 2.3 | 3.80 | 0.53 | 0.38 |

stepwise-regression procedure. Forward stepwise-multiple regression was carried out using the R statistical computing software. For each flow parameter, the forward stepwise method started by considering a simple linear model for predicting the parameter as a constant value (i.e. starting with no variables in the model). The method adds an extra explanatory variable (if any) to the model at each step, choosing the variable that minimizes the Akaike information criterion (*AIC*) which is a measure of the relative quality of a statistical model (Lindsey and Sheather, 2010). As shown in Fig. 7, the forward stepwise procedure is repeated until no further reductions in *AIC* can be obtained. The best set of explanatory variables (catchment characteristics) and the estimates of $\beta_i$, (with $i = 0, 1, \ldots, N$) for the regression models were identified using an ordinary least-squares (OLS) algorithm.

The prediction accuracy of the non- and log-transformed models was evaluated and compared using the adjusted $R^2$ from the MASS package in R software (Ripley et al., 2013) as well as the mean absolute error (*MAE*), root mean square error (*RMSE*) and adjusted Nash-Sutcliffe Efficiency ($NSE_{adj}$,) from the hydroGOF package (Zambrano-Bigiarini, 2014). The hydroGOF pack is oriented for use during calibration, validation, and application of hydrological models. Equations for the evaluation statistics are shown below.

$$R^2 = \frac{\sum_i (\hat{x}_i - \bar{x})^2}{\sum_i (x_i - \bar{x})^2} \tag{4}$$

$$R^2_{adj} = 1 - \frac{(1 - R^2)(N - 1)}{N - (p + 1)} \tag{5}$$

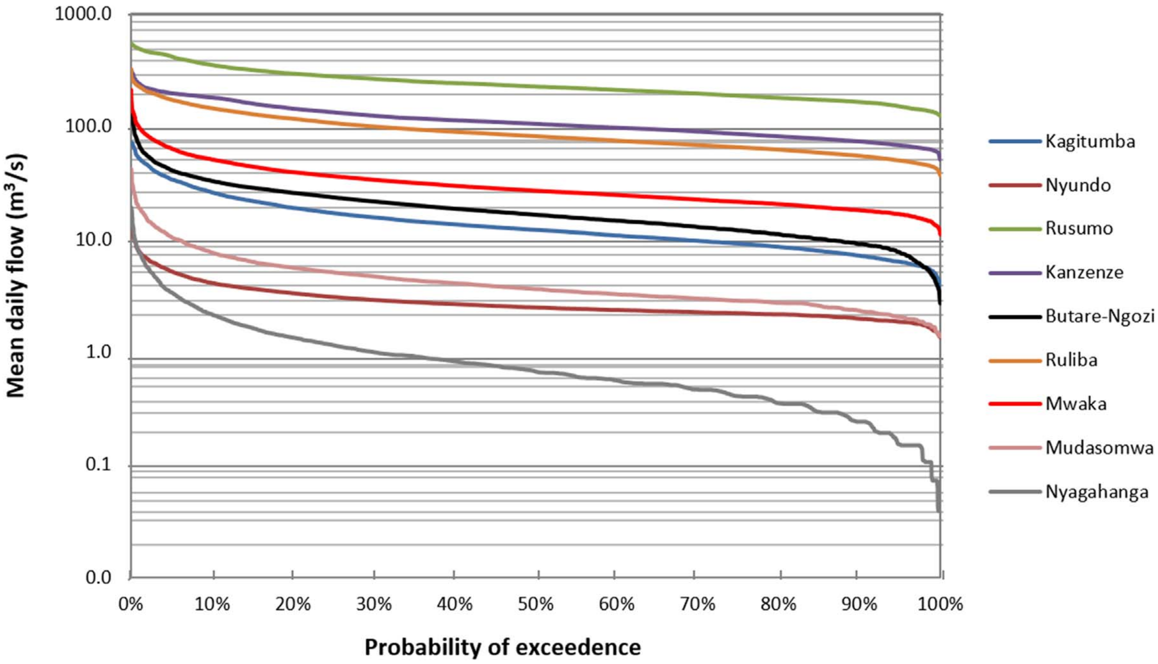


**Fig. 6.** Flow duration curves for the 9 gauged catchments.

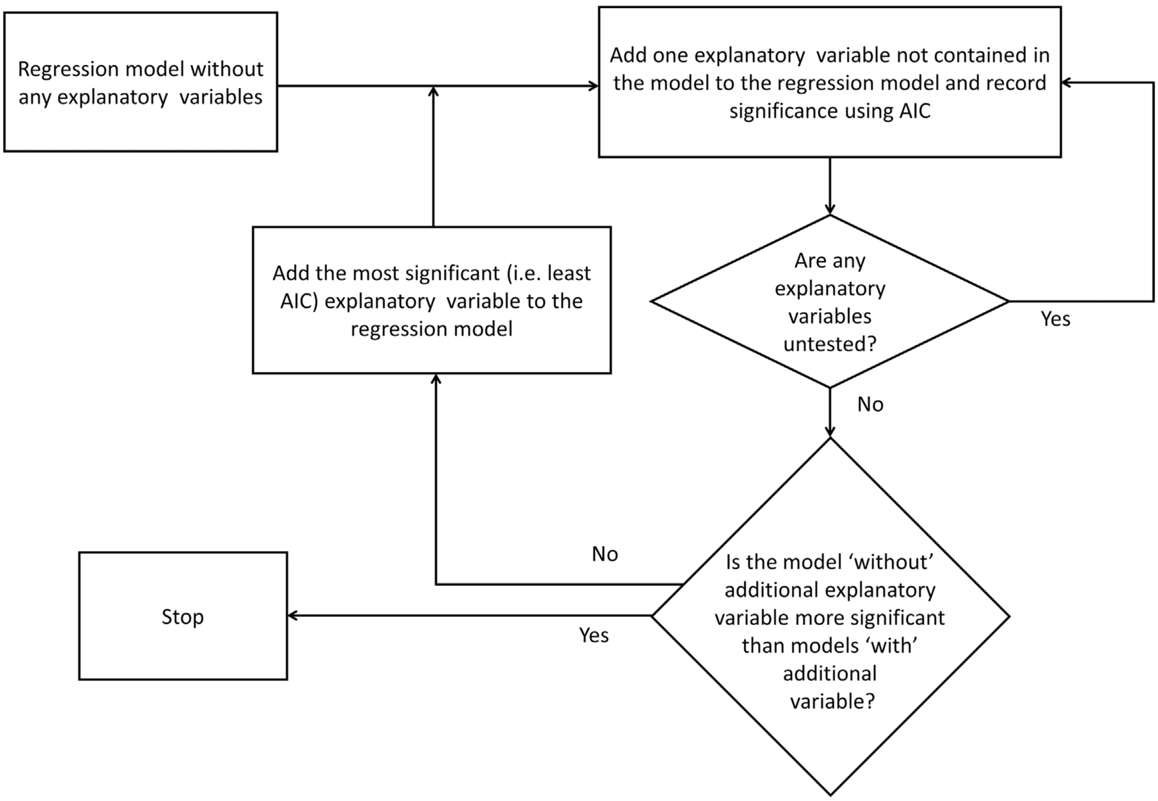


**Fig. 7.** Flowchart of the forward stepwise regression procedure.

$$MAE = \frac{\sum_i |\hat{x}_i - x_i|}{N} \tag{6}$$

$$RMSE = \sqrt{\frac{\sum_i (\hat{x}_i - x_i)^2}{N}} \tag{7}$$

$$NSE = 1 - \frac{\sum_i (x_i - \hat{x}_i)^2}{\sum_i (x_i - \bar{x})^2} \tag{8}$$

$$NSE_{adj} = 1 - \frac{(1 - NSE)(N - 1)}{N - (p + 1)} \tag{9}$$

$x_i$ = observed values
$\hat{x}_i$ = predicted values
$\bar{x}$ = observed mean value
$N$ = number of catchments
$p$ = number of explanatory variables

In order to compare the non- and log-transformed multi-regression models and to select the most suitable models, the performance of the models were examined using the corrected Akaike information criterion ($AIC_C$) and Leave-one-out cross-validation (*LOOCV*). *AIC* is asymptotically equivalent to *LOOCV* when *N* is large (Chen et al., 2013; Simon et al., 2003). Since *N* is much less than the total number of explanatory variables ($p$) considered in this study, $AIC_C$ which is recommended when the number of observations (*N*) is less than 40 times the number of parameters (Burnham and Anderson, 2002) was considered. It is related to the usual *AIC* by

$$AIC_C = AIC + \frac{2p \times (N - 1)}{N - (p + 1)} \tag{10}$$

and

$$AIC = 2p - 2\ln(L) \tag{11}$$

where *L* is the maximized value of the likelihood function for the estimated model. *AIC* values for all models were computed using the MASS package in R software (Ripley et al., 2013).

Leave-one-out cross-validation (*LOOCV*) is the degenerate case of K-Fold cross-validation, where K is chosen as the total number of catchments (*N*). In this approach, each gauged catchment was treated as ungauged and the parameters for that catchment estimated from remaining gauged catchments. The procedure was repeated for each catchment in turn and the *LOOCV* was computed for each flow parameter as the mean square error using R statistical computing software, and in particular the "cv.lm()" function in the DAAG package (Maindonald and Braun 2013).

Although $AIC_C$ penalizes for the number of explanatory variables and small sample size, $AIC_C$ and *LOOCV* do not always lead to the same model being selected for a small sample size. Thus, this study gives preference to *LOOCV* in model selection because it considers all the uncertainty components which include input data uncertainty, model uncertainty and parameter uncertainty (Wagener and Montanari, 2011). In a case where two or more models have almost equal *LOOCV* values, the model with the lowest *AICc* amongst them is selected if it also has the lowest number of explanatory variables.

Evaluating the significance of the explanatory variables identified for predicting the different flow parameters requires some *a priori* knowledge of the physical relationships between physiographic and climatic variables. Regrettably, these relationships are not well defined at catchment scale. Although forward stepwise regression analysis can identify the physiographic and climatic variables that are good predictors of the flow parameters, as stated by Sefton and Howarth (1998), the relationships identified are only empirical. Some of these relationships may simply be an accident of the data although statistically significant. While some relationships may have strong physical significance, others may be surrogate predictors or may represent process interactions which cannot be explained by current knowledge of the relationships between hydrological processes and parameters controlling them at the catchment scale (Mohamoud, 2008).

## 3. Results and discussions

Table 5 shows the explanatory variables for the best non-transformed regression models for each flow parameter, along with residual standard error, $R^2_{adj}$, $NSE_{adj}$, *MAE*, *RMSE*, $AIC_C$ and the *LOOCV*. It should be noted that even though models with more

**Table 5**
Non-transformed regression models.

| Parameter | Explanatory variables | Residual standard error | $R^2_{adj}$ | $NSE_{adj}$ | *MAE* | *RMSE* | *AICc* | *LOOCV* |
|---|---|---|---|---|---|---|---|---|
| $q_5$ | $L_F$ | 4.38 | 0.81 | 0.81 | 3.35 | 3.86 | 28.9 | 23.6 |
| $q_{mean}$ | $PREC,\ L_F$ | 1.22 | 0.93 | 0.93 | 0.84 | 1.00 | 7.98 | 3.8 |
| $q_{50}$ | $PREC,\ L_F,\ L_{HP}$ | 1.03 | 0.93 | 0.93 | 0.62 | 0.77 | 8.08 | 1.8 |
| $q_{95}$ | $PREC$ | 1.25 | 0.78 | 0.78 | 0.93 | 1.11 | 6.37 | 2.2 |
| $FDC_{slp}$ | $H_+,\ L_C,\ L_G,\ H_m$ | 0.02 | 0.95 | 0.93 | 0.01 | 0.01 | −59.6 | 0.0 |

**Table 6**
Log-transformed regression models.

| Parameter | Explanatory variables | Residual Standard Error | $R^2_{adj}$ | $NSE_{adj}$ | $MAE$ | $RMSE$ | $AICc$ | $LOOCV$ |
|---|---|---|---|---|---|---|---|---|
| $q_5$ | $PREC$, $D$ | 0.08 | 0.87 | 0.81 | 2.80 | 3.58 | −41.91 | 1.0 |
| $q_{mean}$ | $PREC$, $L_C$ | 0.06 | 0.94 | 0.88 | 1.04 | 1.35 | −47.03 | 1.0 |
| $q_{50}$ | $PREC$, $C_S$ | 0.06 | 0.94 | 0.87 | 0.85 | 1.18 | −45.08 | 1.0 |
| $q_{95}$ | $PREC$, $C_S$, $ET_{pot}$, $H_R$ | 0.07 | 0.97 | 0.89 | 0.46 | 0.58 | −36.28 | 1.0 |
| $FDC_{slp}$ | $H_+$, $L_{HP}$, $ET_{pot}$ | 0.05 | 0.87 | 0.80 | 0.02 | 0.03 | −45.05 | 1.0 |

explanatory variables generally tend to have better fit between predicted and observed values (in terms of higher $R^2_{adj}$ and $NSE_{adj}$, and lower standard residual error, $MAE$ and $RMSE$), this does not necessarily imply that the addition of more variables improves the predictive performance of models. Although $R^2_{adj}$ and $NSE_{adj}$ are widely used in most studies, their tendency to select models with too many variables makes them less suitable for prediction than either $AIC_C$ or $LOOCV$.

The best non-transformed stepwise-regression model for each flow parameter was selected on the basis of $LOOCV$ except in a case whereby two or more models have very similar $LOOCV$ values. In this case, the model with the lowest $AICc$ amongst them is selected if it also has the lowest number of explanatory variables. For $q_{mean}$, $q_{95}$ and $FDC_{slp}$, the $AIC_C$ and $LOOCV$ indicated the same 'best' models.

For predicting high flows ($q_5$), it was observed that percent of forested area ($L_F$) is the only explanatory variable. For median and mean flows, it was observed that precipitation ($PREC$) and $L_F$ are the two dominant explanatory variables. For predicting low flows ($q_{95}$), only $PREC$ appears in the stepwise model. The significant explanatory variables for estimating the slope of flow duration curve ($FDC_{slp}$) using non-transformed model are maximum elevation ($H_+$), percent cropland ($L_C$), percent grassland ($L_C$) and mean elevation ($H_m$).

The non-transformed OLS regression models developed were based on the assumptions that the residuals (predicted minus observed) are independent, homoscedastic and normally distributed. Since non-transformed linear regression models often exhibit heteroscedasticity (Viglione et al., 2007; Vezza et al., 2010), logarithmic transformations were used on all flow parameters and explanatory variables (except land use variables which have some zero values) to avoid heteroscedasticity and non-normality of the residuals of the regressions.

Table 6 shows the explanatory variables for the best log-transformed regression models along with residual standard error, $R^2_{adj}$, $NSE_{adj}$, $MAE$, $RMSE$, $AIC_C$ and the $LOOCV$. It should be noted that except for $FDC_{slp}$, the $AIC_C$ and $LOOCV$ indicated the same "best" models for the flow quantiles. The best log-transformed stepwise regression models for all flow parameters were chosen based on $LOOCV$ with the exception of $FDC_{slp}$ for which $AIC_C$ was used since the model with 3 explanatory variables (least $AIC_C$) and that with 5 variables (least $LOOCV$) both had almost equal $AIC_C$ and $LOOCV$ values. The removal of two explanatory variables from the $LOOCV$ model simplified the model without significantly harming the predictive ability of the $AIC_C$ model.

It was observed that all flow quantiles had precipitation ($PREC$) as the dominant explanatory variable while elevation characteristics had the most significance for $FDC_{slp}$ prediction. Hope and Bart (2012) also found precipitation to be the dominant predictor for similar flow quantiles in a study in Cape Floristic Region of South Africa.

The performances (in terms of $LOOCV$) of non- and log-transformed models were compared to make the final selection of the overall best-performing regression models. In general, $LOOCV$ values were much smaller for log-transformed models than for non-transformed models. The stepwise regression analysis identified precipitation ($PREC$) and river density ($D$) as the overall best predictors for $q_5$. The selection of $D$ as the second explanatory variable indicates that high flow response of the catchments is also controlled by the flow concentration processes and runoff. This seems reasonable as the negative relationship between $q_5$ and $D$ ($r = -0.45$) can be explained by the fact that specific high flow discharges $q_5$ (l/(s km$^2$)) were computed by standardizing $Q_5$ values by respective catchment areas in order to make the values more comparable across scales. As expected, the correlation between $Q_5$ and $D$ is positive ($r = 0.63$) thus justifying the high $D$ (which depends on soil permeability and underlying rock type) and high $Q_5$ found in rugged regions and catchments with high range elevation. In a similar study by Laio et al. (2011), the power law regression model for predicting peak flows also included precipitation and a soil permeability index.

For $q_{mean}$, the log-transformed model also out-performed the non-transformed model. Among the response descriptors, precipitation ($PREC$) and percentage of cropland ($L_C$) had the most significant effects on $q_{mean}$ and were selected as the best explanatory variables. High precipitation produces high $q_{mean}$. Precipitation has also been used previously in studies by Wolock et al. (2004) and Santhi et al. (2007). The relevance of $L_C$ can be justified by the fact that cropland, which is the largest land cover in the gauged catchments, has higher runoff coefficient when compared with forest, shrubland and grassland. Increasing $L_C$ will increase $Q_{mean}$ but will reduce the specific mean flow ($q_{mean}$) due to the negative correlation of $q_{mean}$ with catchment area.

The log-transformed model gave the best $LOOCV$ value for $q_{50}$, and precipitation ($PREC$) was the best explanatory variable, followed by mean catchment slope ($C_S$). The exponent related to $C_S$ is $-0.93$, meaning that $q_{50}$ actually decreases with increasing mean catchment slope ($r = -0.22$). Unsurprisingly, $Q_{50}$ is also negatively correlated to $C_S$ ($r = -0.48$). There is less storage in small catchments with high slopes due to high surface runoff and rapid drop in water table (Rastogi, 1988; Paniconi et al., 2003) which results in quick drainage of the catchments. Since smaller catchments tend to have larger $C_S$ in Rwanda, an increase in $C_S$ has a negative effect on $q_{50}$.

The best model for predicting $q_{95}$ was the log-transformed model. Low flow conditions were influenced by both the climate and physiography of the catchments. Precipitation (*PREC*), catchment slope ($C_S$), potential evapotranspiration ($ET_{pot}$) and range elevation ($H_R$) were identified as the significant variables controlling water fluxes at the land surfaces of the catchments in dry seasons when there is little precipitation and high evaporation rates. Catchment and climatic characteristics such as slope and mean annual precipitation were also used in similar studies for low-flow analysis (Flynn, 2003; Wright and Ensminger, 2004).

The positive relationship between precipitation and low flow in the model is obvious. Water which is stored in various ways in the catchments (in groundwater systems, in soil, in lakes, etc.) is released at different times to the streams and rivers. Given that smaller gauged catchments in Rwanda tend to have larger $C_S$, an increase in $C_S$ has a negative effect on $q_{95}$. A study by Pearson (1995) also showed a negative relationship between log specific discharge and catchment slope. As expected, the negative exponent of evapotranspiration indicates loss of water from catchments when it increases. The effect of a combination of range elevation ($H_R$) and catchment slope ($C_S$) is capable of describing a catchment's shape and size, which both have effect on the streamflow regime and thereby playing a significant role on $q_{95}$ prediction. Based on the signs of the exponents, the ratio $H_R/C_S$ may be seen as a measure of the main channel length of a catchment. A study in Italy (Castellarin et al., 2004) with climate and topography relatively similar to Rwanda shows that relief and main channel length play significant roles in low flow predictions.

The best model selected for predicting slope of flow duration curve ($FDC_{slp}$) was the log-transformed model. Maximum elevation ($H_+$), percentage of hardpans ($L_{HP}$), and potential evapotranspiration ($ET_{pot}$) are identified to be the most significant variables for the regionalization of $FDC_{slp}$. According to Castellarin et al. (2013), current studies on flow duration curves predictions depend considerably on statistical methods, and topographic elevation (in mountainous regions) is among the most frequently used predictors of the slope and shape of FDCs. In this study, it was observed that maximum elevation ($H_+$) is positively correlated to area ($A$), and that both $H_+$ and $A$ are negatively correlated to $FDC_{slp}$. This has a significant influence on catchment response time since larger catchments flow for more of the time. Thus the relevance of $H_+$ can be justified if one assumes that the slope of the FDC probably decreases with the area of the catchment as a result of larger storage capacities. As expected, the regression model indicates higher $FDC_{slp}$ for catchments with larger percentage of impermeable salt hardpans ($L_{HP}$) since there will be less infiltration and release of water to streams and rivers during low-flow periods. Catchments with high percentage of hardpans ($L_{HP}$) are dominated by overland flow and are usually quickly responding or 'flashy' catchments. The positive exponent of potential evapotranspiration ($ET_{pot}$) can be explained by the FDC response to vegetation. An increase in the percentage of forest or shrubland in a catchment may lead to reduced high flows and even more reduced low flows (i.e. higher $FDC_{slp}$). Similar results were found by Schofield (1996) and Vertessy (2000) that deforestation in some Australian catchments led to a quick rise of the groundwater table, and associated groundwater flow which resulted in large increases in low flows.

The developed models are limited to the study area and to the ranges in the values of the gauged catchment characteristics used in estimating the coefficients of regression. Owing to the high spatio-temporal heterogeneity of topography and land cover characteristics in Rwanda, using the models to estimate flow parameters of ungauged catchments remains full of uncertainties as the accuracy of the estimates might decrease. For example, the percentage of forest ($L_F$) of an ungauged catchment may be very large compared to the $L_F$ for the gauged catchments used in the regression analysis. If $L_F$ is not selected as an explanatory variable in the stepwise procedure, the regression model will not detect the effect of extrapolation when transferring the information to the ungauged catchment. Hence care should be taken by evaluating the ungauged catchments of interest to find out if the regression models are appropriate for their intended use. In addition, in order to apply the procedure to other regions around Rwanda with similar climate and landscape, new regression models have to be developed. Better quality data and more well-gauged stations might give different estimates with less uncertainty (Snelder et al., 2013).

### 3.1. Spatial patterns of predictions of ungauged catchments

The spatial patterns of flow parameter predictions in ungauged catchments were assessed by maps (Fig. 8) in the study area. The figure uses the same intervals for comparison of the flow quantiles. Each panel shows a flow parameter predicted using the overall "best" regional regression model. For $q_5$, $q_{mean}$ and $q_{50}$, the predicted values are large in the west of the country and near the Congo-Nile drainage divide southwest of the country while values are low in the eastern regions. To some extent, the spatial patterns for $q_5$, $q_{mean}$ and $q_{50}$ look similar to the spatial distribution of mean annual precipitation (Fig. 2). This is because, in general, and being the most significant explanatory variable, precipitation increases westward to 1400–1500 mm near the Congo-Nile drainage divide. It reaches more than 1500 mm in the southwestern part, up to 2000–2400 mm in the region of the northwestern volcano peaks, and about 1000 mm or less in the eastern regions.

Although the spatial patterns for $q_{mean}$ and $q_{50}$ are quite similar, they differ from each other due to the different second explanatory variables selected by their respective models (Table 6). Predicted low flow ($q_{95}$) values are lower than 3.23 L/(s km$^2$) for most of the catchments in the Nile Basin when compared with those in the Congo Basin. During dry seasons, streamflow is greatly reduced by the evapotranspiration from the forest, cropland and shrubland which consume water from the soil and the groundwater system thus reducing baseflow. The spatial pattern for $q_{95}$ shows agriculture and hydropower production could be affected more in the eastern parts. Catchments in the eastern regions are farther from the drainage divide which reduces the effect of regional sub-surface flow systems contributing to baseflow. The regional sub-surface flow from high precipitation regions near the divide will likely end up in Lake Victoria farther east of Rwanda.

Relatively high $FDC_{slp}$ predictions in the eastern catchments indicate highly variable streamflow mainly from direct runoff from open shrublands. $FDC_{slp}$ values are low in the western catchments close to the Congo-Nile drainage divide which denotes the presence of surface- or ground-water storage as a result of forest which increases infiltration and leads to reduced high flows.

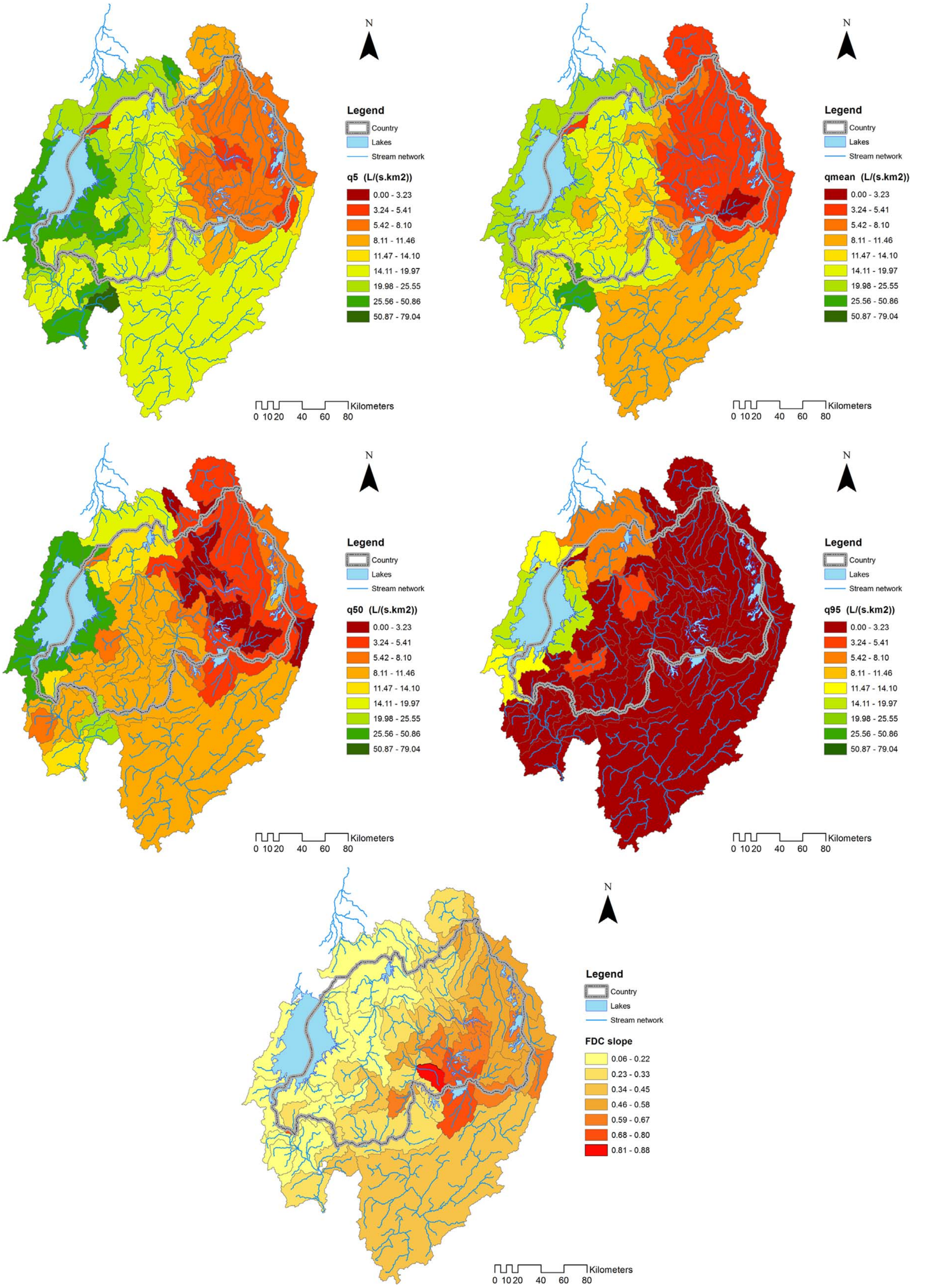


**Fig. 8.** Spatial patterns of $q_5$, $q_{mean}$, $q_{50}$, $q_{95}$ and $FDC_{slp}$ predictions for ungauged catchments.

## 4. Conclusions

This study presented a method to regionalize flow parameters in ungauged catchments in Rwanda. A combination of climate similarity and parameter regression was investigated as a method to develop models for predictions at ungauged catchments. The regionalized flow parameters were $q_5$, $q_{mean}$, $q_{50}$, $q_{95}$ and $FDC_{slp}$. A regression model was developed for each parameter using a forward stepwise-regression procedure and by considering non-transformed and log-transformed equations. Since the aim of the study was to predict the parameters in ungauged catchments, leave-one-out cross validation was used as the criterion for selecting the best performing models.

The results indicated that the log-transformed models out-performed the non-transformed ones. Using the log-transformed models, climate, physiographic and land cover descriptors were observed to strongly influence the hydrology of the study area. The most significant climate descriptor for all flow quantiles was mean annual precipitation. River density, mean catchment slope, minimum and maximum catchment elevation were among the dominant physiographical descriptors for the flow parameters. The dominant land cover descriptors identified were percentages of cropland and shrubland salt hardpans.

It is believed that the method presented in this study is a valuable tool for the prediction of flow parameters in ungauged catchments. The models developed can be used by government agencies and catchment stakeholders for water resources assessments and for better understanding of water fluxes in Rwandan catchments. Although the models were developed using data for the study area, parameter regression can be used anywhere in the world if regional regression models are established following the approach presented in this study. This depends on knowledge of the catchment, data quality and availability, and a thorough understanding of all the explanatory variables used.

## Conflict of interest

None.

## References


Blöschl, G., Sivapalan, M., 1995. Scale issues in hydrological modelling: a review. Hydrol. Process. 9, 251–290.

Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., Savenije, H., 2013. Runoff Prediction in Ungauged Basins: Synthesis Across Processes, Places and Scales. Cambridge University Press, UK, pp. 252.

Blöschl, G., 2005. Rainfall-runoff modeling of ungauged catchments. In: Anderson, M.G. (Ed.), Encyclopedia of Hydrological Sciences. John Wiley & Sons Chichester, pp. 2061–2080 Article 133.

Budyko, M.I., 1974. Climate and Life. Academic Press, New York.

Burn, D.H., Boorman, D.B., 1993. Estimation of hydrological parameters at ungauged catchments. J. Hydrol. 143, 429–454. http://dx.doi.org/10.1016/0022-1694(93)90203-L.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Interference: A Practical HydrInformation-Theoretic Approach. Springer, New York 353 pp.

Castellarin, A., Galeati, G., Brandimarte, L., Montanari, A., Brath, A., 2004. Regional flow-duration curves: reliability for ungauged basins. Adv. Water Resour. 27, 953–965. http://dx.doi.org/10.1016/j.advwatres.2004.08.005.

Castellarin, A., Botter, G., Hughes, D.A., Liu, S., Ouarda, T.B., Parajka, J., Post, D.A., Sivapalan, M., Spence, C., Viglione, A., Vogel, R.M., 2013. Prediction of Flow Duration Curves in Ungauged Basins, 1st edition. Cambridge University Press, Cambridge, pp. 135–162. http://dx.doi.org/10.1017/CBO9781139235761.010.

Chen, A., Owen, A.B., Shi, M., 2013. Data Enriched Linear Regression. arXiv:1304.1837, 6 Apr 2013. Accessed from http://arxiv.org/pdf/1304.1837.pdf on 17/02/2014.

Flynn, R.H., 2003. Development of regression equations to estimate flow durations and Low-flow-frequency statistics in New Hampshire streams. U.S. Geological Survey Water-Resources Investigations Report 02-4298.

GOR, 2007. Economic Development and Poverty Reduction Strategy 2008–2012. The Republic of Rwanda http://siteresources.worldbank.org/INTRWANDA/Resources/EDPRS-English.pdf.

Gerrits, A.M., Savenije, H.H., Veling, E.J., Pfister, L., 2009. Analytical derivation of the Budyko curve based on rainfall characteristics and a simple evaporation model. Water Resour. Res. 45http://dx.doi.org/10.1029/2008WR007308. W04403.

He, Y., Bardossy, A., Zehe, E., 2011. A review of regionalization for continuous streamflow simulation. Hydrol. Earth Syst. Sci. 15, 3539–3553. http://dx.doi.org/10.5194/hess-15-3539-2011.

Hope, A., Bart, R., 2012. Synthetic monthly flow duration curves for the Cape Floristic Region, South Africa. Water SA 38 (2). http://dx.doi.org/10.4314/wsa.v38i2.4.

Hrachowitz, M., Savenije, H.H.G., Blöschl, G., McDonnell, J.J., Sivapalan, M., Pomeroy, J.W., Arheimer, B., Blume, T., Clark, M.P., Ehret, U., Fenicia, F., Freer, J.E., Gelfan, A., Gupta, H.V., Hughes, D.A., Hut, R.W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P.A., Uhlenbrook, S., Wagener, T., Winsemius, H.C., Woods, R.A., Zehe, E., Cudennec, C., 2013. A decade of predictions in ungauged basins (PUB)—a review. Hydrol. Sci. J. 58 (6), 1198–1255. http://dx.doi.org/10.1080/02626667.2013.803183.

Kokkonen, T.S., Jakeman, A.J., Young, P.C., Koivusalo, H.J., 2003. Predicting daily flows in ungauged catchments: model regionalization from catchment descriptors at the Coweeta Hydrologic Laboratory, North Carolina. Hydrol. Process. 17, 2219–2238. http://dx.doi.org/10.1002/hyp.1329.

Laaha, G., Blöschl, G., 2006. Seasonality indices for regionalizing low flows. Hydrol. Process. 20, 3851–3878. http://dx.doi.org/10.1002/hyp.6161.

Laaha, G., Blöschl, G., 2007. A national low flow estimation procedure for Austria. Hydrol. Sci. J. 52 (4), 625–644. http://dx.doi.org/10.1623/hysj.52.4.625. ISSN: 02626667.

Laio, F., Ganora, D., Claps, P., Galeati, G., 2011. Spatially smooth regional estimation of the flood frequency curve (with uncertainty). J. Hydrol. 408, 67–77. http://dx.doi.org/10.1016/j.jhydrol.2011.07.022.

Li, H., Zhang, Y., Chiew, F.H.S., Xu, S., 2009. Predicting runoff in ungauged catchments by using Xinanjiang model with MODIS leaf area index. J. Hydrol. 370, 155–162. http://dx.doi.org/10.1016/j.jhydrol.2009.03.003.

Lindsey, C., Sheather, S., 2010. Variable selection in linear regression. Stata J. 10 (4), 650–669.

MINECOFIN, 2000. Rwanda Vision 2020. Ministry of Finance and Economic Planning, Republic of Rwanda.

MININFRA, 2011. Annual Report 2010. Ministry of Infrastructure, Republic of Rwanda.

Maindonald, J.H., Braun, W.J., 2013. Package 'DAAG'. Data Analysis and Graphics Data and Functions Version 1.18. Accessed from http://cran.r-project.org/web/packages/DAAG/DAAG.pdf on 02/01/2014.

Mazvimavi, D., Meijerink, A.M.J., Stein, A., 2004. Prediction of base flows from basin characteristics: a case study from Zimbabwe. Hydrol. Sci J. 49 (4), 703–715.

Mazvimavi, D., Meijerink, A.M.J., Savenije, H.H.G., Stein, A., 2005. Prediction of flow characteristics using multiple regression and neural networks: a case study in Zimbabwe. Phys. Chem. Earth 30, 639–647.

McIntyre, N.R., Lee, H., Wheater, H.S., Young, A.R., 2004. Tools and approaches for evaluating uncertainty in streamflow predictions in ungauged UK catchments, Complexity and Integrated Resources Management. In: Pahl-Wostl, C., Schmidt, S., Jakeman, T. (Eds.), Proc. IEMSS International Congress. Osnabrueck, Germany.
McNulty, S., Mack, E.C., Sun, G., Caldwell, P., 2016. Hydrologic Modeling for Water Resource Assessment in a Developing Country: The Rwanda Case Study. pp. 183 Chapter 2.6.
McSweeney, R., 2011. Rwanda's Climate: Observations and Projections – Appendix E. Smith School of Enterprise and the Environment, University of Oxford. Accessed from http://www.smithschool.ox.ac.uk/research/library/Rwanda_Climate-data-final_proofed.pdf on 02/10/2013.
Merz, R., Blöschl, G., 2004. Regionalisation of catchment model parameters. J. Hydrol. 287, 95–123. http://dx.doi.org/10.1016/j.jhydrol.2003.09.028. ISSN: 00221694.
Mohamoud, Y.M., 2008. Prediction of daily flow duration curves and streamflow for ungauged catchments using regional flow duration curves. Hydrol. Sci. J. 53 (4), 706–724. http://dx.doi.org/10.1623/hysj.53.4.706.
Museruka, C., Ndayisaba, C., Mutabazi, A., Nsabukunze, F., 2011. Assessment of temperature changes in Upper Akagera basin: case of Rwanda. Nile Basin Water Sci. Eng. J. 4 (2), 47–58.
Oudin, L., Kay, A., Andreassian, V., Perrin, C., 2010. Are seemingly physically similar catchments truly hydrologically similar? Water Resour. Res. 46http://dx.doi.org/10.1029/2009WR008887. W11558.
Paniconi, C., Troch, P.A., van Loon, E.E., Hilberts, A.G., 2003. Hillslope-storage Boussinesq model for subsurface flow and variable source areas along complex hillslopes: 2. Intercomparison with a three-dimensional Richards equation model. Water Resour. Res. 39 (11), 1317. http://dx.doi.org/10.1029/2002wr001730.
Parajka, J., Merz, R., Blöschl, G., 2005. A comparison of regionalization methods for catchment model parameters. Hydrol. Earth Syst. Sci. 9, 157–171. http://dx.doi.org/10.5194/hess-9-157-2005.
Parajka, J., Viglione, A., Rogger, M., Salinas, J.L., Sivapalan, M., Blöschl, G., 2013. Comparative assessment of predictions in ungauged basins – part 1: runoff-hydrograph studies. Hydrol. Earth Syst. Sci. 17, 1783–1795. http://dx.doi.org/10.5194/hess-17-1783-2013.
Pearson, C.P., 1995. Regional flow frequency analysis of low flows in New Zealand rivers. J. Hydrol. (NZ) 33 (2), 94–122.
Pike, J.G., 1964. The estimation of annual runoff from meteorological data in a tropical. Clim. J. Hydrol. 2, 116–123.
RIWSP, 2012a. In: Venneker, R., Wenninger, J. (Eds.), RNRA Hydrometric Observing Network – Observing Station Details, Version 1. Internal report to the Rwanda Natural Resources Authority, RIWSP-No. 2.2.1/05 Rwanda Integrated Water Security Program, Kigali, Rwanda.
RIWSP, 2012b. RWRIS HydroScape Version 2.0. In: Wenninger, J., Venneker, R. (Eds.), Rwanda Integrated Water Security Program/Rwanda Natural Resources Authority, Kigali.
RIWSP, 2012c. In: Venneker, R., Wenninger, J. (Eds.), Review of the National Hydrological Service in Rwanda. RIWSP-No. 2.2.1/02 Rwanda Integrated Water Security Program, Kigali, Rwanda.
RIWSP, 2012d. In: Venneker, R., Wenninger, J. (Eds.), RNRA Hydrometric Archive – Data Conversion and Recovery Report, Version 2. Internal Report to the Rwanda Natural Resources Authority, RIWSP-No. 2.2.1/04 Rwanda Integrated Water Security Program, Kigali, Rwanda.
Randrianasolo, A., Ramos, M.H., Andreassian, V., 2011. Hydrological ensemble forecasting at ungauged basins: using neighbor catchments for model setup and updating. Adv. Geosci. 29, 1–11. http://dx.doi.org/10.5194/adgeo-29-1-2011.
Rastogi, A.K., 1988. Flow studies through porous and permeable sub-surface water bearing bodies having sloping base. Irrig. Power 45 (1), 84–90.
Ripley, B., Venables, B., Bates, D.M., Hornik, K., Gebhardt, A., Firth, D., 2013. P. Ackage 'MASS'. Support Functions and Datasets for Venables and Ripley's MASS. . Accessed from http://www.stats.ox.ac.uk/pub/MASS4/ on 05/12/2013.
Santhi, C., Allen, P.M., Muttiah, R.S., Arnold, J.G., Tuppad, P., 2007. Regional estimation of base flow for the conterminous United States by hydrologic landscape regions. J. Hydrol. 351, 139–153. http://dx.doi.org/10.1016/j.jhydrol.2007.12.018.
Schofield, N.J., 1996. Forest management impacts on water values. Recent Res. Dev. Hydrol. 1, 1–20.
Schreiber, P., 1904. Uber die Beziehungen zwischen dem Niederschlag und der Wasserfuhrung der Flusse in Mitteleuropa. Z. Meteorol. 21 (10), 441–452.
Sefton, C.E.M., Howarth, S.M., 1998. Relationships between dynamic response characteristics and physical descriptors of catchments in England and Wales. J. Hydrol. 211, 1–16. http://dx.doi.org/10.1016/S0022-1694(98)00163-2.
Seibert, J., Beven, K.J., 2009. Gauging the ungauged basin: how many discharge measurements are needed? Hydrol. Earth Syst. Sci. 13, 883–892. http://dx.doi.org/10.5194/hess-13-883-2009.
Simon, G., Lendasse, A., Wertz, V., Verleysen, M., 2003. In: ESANN 2003?European Symposium on Artificial Neural Networks. Springer–Verlag, Berlin, Heidelbergg.
Sivapalan, M., Takeuchi, K., Franks, S., Gupta, V., Karambiri, H., Lakshmi, V., Liang, X., Mcdonnell, J., Mendiondo, E., O'connell, P., Oki, T., Pomeroy, J., Schertzer, D., Uhlenbrook, S., Zehe, E., 2003. IAHS decade on predictions in Ungauged Basins (PUB), 2003–2012: shaping an exciting future for the hydrological sciences. Hydrolog. Sci. J. 48 (6), 857–880. http://dx.doi.org/10.1623/Hysj.48.6.857.51421.
Skoien, J.O., Blöschl, G., 2007. Spatiotemporal topological kriging of runoff time series. Water Resour. Res. 43, W09419. http://dx.doi.org/10.1029/2006WR005760.
Skoien, J.O., Merz, R., Blöschl, G., 2006. Top-kriging – geostatistics on stream networks. Hydrol. Earth Syst. Sci. 10, 277–287. http://dx.doi.org/10.5194/hess-10-277-2006.
Snelder, T.H., Datry, T., Lamouroux, N., Larned, S.T., Sauquet, E., Pella, H., Catalogne, C., 2013. Regionalization of patterns of flow intermittence from gauging station records. Hydrol. Earth Syst. Sci. 17, 2685–2699. http://dx.doi.org/10.5194/hess-17-2685-2013.
Vertessy, R.A., 2000. Impacts of plantation forestry on catchment runoff. In: Sadanandan Nabia, E.K., Brown, A.G. (eds). Plantations, farm forestry and water. Proceedings of a National Workshop, 20–21 July, Melbourne, 9–19.
Vezza, P., Comoglio, C., Rosso, M., Viglione, A., 2010. Low flows regionalization in North-Western Italy. Water Resour. Manage. 24, 4049–4074. http://dx.doi.org/10.1007/s11269-010-9647-3.
Viglione, A., Claps, P., Laio, F., 2007. Mean annual runoff estimation in North-Western Italy. In: La Loggia, G., Aronica, G.T., Ciraolo, G. (Eds.), Water Resources Assessment and Management Under Water Scarcity Scenarios. CDSU Publ., Milano, pp. 97–122.
WFP, 2009. Rwanda: comprehensive food security and vulnerability analysis and nutrition survey July 2009 (data collected in February–March 2009). World Food Program, VAM Food Security Analysis.
Wagener, T., Montanari, A., 2011. Convergence of approaches toward reducing uncertainty in predictions in ungauged basins. Water Resour. Res. 47. http://dx.doi.org/10.1029/2010wr009469.
Wagener, T., Wheater, H.S., 2006. Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty. J. Hydrol. 320, 132–154.
Wolock, D.M., Winter, T.C., McMohan, G., 2004. Delineation and evaluation of hydrologic-landscape regions in the United States using geographic information system tools and multivariate statistical analyses. Environ. Manage. 34 (1), 71–88.
Wright, L.S., Ensminger, P.A., 2004. Regionalized regression equations for estimating low-flow characteristics for selected Louisiana streams. U.S. Geological Survey Water Resources Technical Report No. 75. . Accessed from http://la.water.usgs.gov/publications/pdfs/TR75.pdf on 19/03/2014.
Young, A.R., 2006. Stream flow simulation within UK ungauged catchments using a daily rainfall-runoff model. J. Hydrol. 320, 155–172.
Zambrano-Bigiarini, M., 2014. Package 'hydroGOF'. Goodness-of-fit Functions for Comparison of Simulated and Observed Hydrological Time Series. Accessed from http://cran.r-project.org/web/packages/hydroGOF/hydroGOF.pdf on 02/01/2014.
Zhang, Y.Q., Chiew, F.H.S., 2009. Relative merits of different methods for runoff predictions in ungauged catchments. Water Resour. Res. 45 W07412, 07410.01029/02008WR007504.