



Classification and flow prediction in a data-scarce watershed of the equatorial Nile region

J.-M. Kileshye Onema^{1,2}, A. E. Taigbenu¹, and J. Ndiritu¹

¹Water Research Group, School of Civil and Environmental Engineering, University of Witwatersrand, Private Bag 3, WITS 2050, South Africa

²WaterNet, Box MP 600 Mount Pleasant, Harare, Zimbabwe

Correspondence to: J.-M. Kileshye Onema (jmkilo3@yahoo.co.uk)

Received: 31 March 2011 – Published in Hydrol. Earth Syst. Sci. Discuss.: 14 April 2011

Revised: 14 April 2012 – Accepted: 21 April 2012 – Published: 15 May 2012

Abstract. Continuous developments and investigations in flow predictions are of interest in watershed hydrology especially where watercourses are poorly gauged and data are scarce like in most parts of Africa. Thus, this paper reports on two approaches to generate local monthly runoff of the data-scarce Semliki watershed. The Semliki River is part of the upper drainage of the Albert Nile. With an average annual local runoff of 4.622 km³/annum, the Semliki watershed contributes up to 20% of the flows of the White Nile. The watershed was sub-divided in 21 sub-catchments (S3 to S23). Using eight physiographic and meteorological variables, generated from remotely sensed acquired datasets and limited catchment data, monthly runoffs were estimated. One ordination technique, the Principal Component Analysis (PCA), and the tree cluster analysis of the landform attributes were performed to study the data structure and spot physiographic similarities between sub-catchments. The PCA revealed the existence of two major groups of sub-catchments – flat (Group I) and hilly (Group II). Linear and nonlinear regression models were used to predict the long-term monthly mean discharges for the two groups of sub-catchments, and their performance evaluated by the Nash-Sutcliffe Efficiency (NSE), Percent bias (PBIAS) and root mean square error to the standard deviation ratio (RSR). The dimensionless indices used for model evaluation indicate that the non-linear model provides better prediction of the flows than the linear one.

1 Introduction

Numerous approaches exist for streamflow prediction in natural river reaches. Streamflow forecasting has significant interest both from a research and an operational point of view. The choice of methods depends on data availability and the type of application. While continuous research efforts strive at enhancing our predictive capability for streamflow, we are often faced with the challenge of making such predictions in basins that are poorly gauged or not gauged at all (Sivapalan et al., 2003). Reliable and accurate estimates of hydrologic components are not only important for water resources planning and management but are also increasingly relevant to environmental studies (Schröder, 2006). Several studies have reported on the use of catchment descriptors and regionalization of parameters for flow prediction in ungauged basins. Among the most recent studies are those of Sefton and Howard (1998), Mwakalila (2003), Xu (2003), Merz and Blöschl (2004), McIntyre et al. (2005), Sanborn and Bledsoe (2006), Yadav et al. (2007), Sharda et al. (2008) Kwon et al. (2009) and Shao et al. (2009). In their comparison of linear regression with artificial neural networks, Heuvelmans et al. (2006) indicated the need for well-informed choice of physical catchment descriptors as a first condition for successful parameter regionalization. Cheng et al. (2006) reported on the importance and usefulness of parsimonious models for runoff prediction in data-poor environment as these models are characterized by few numbers of parameters. Many authors have also identified the reduction of uncertainty associated with predictions in ungauged basins as being very important (Uhlenbrook and Siebert, 2005; Koutsoyiannis, 2005a, b; Zhang et al., 2008).

Table 1. Statistics of physiographic variables.

Variables	No.	Mean	Standard Deviation
Stream Length (x_1)	21	29.46	20.66
Drainage density (x_2)	21	7.85×10^{-2}	3.52×10^{-2}
Average stream Slope (x_3)	21	0.2	0.23
Maximum elevation (x_4)	21	2577.9	1273.84
Minimum elevation (x_5)	21	733.18	99.78
Weighted average elevation (x_6)	21	1164.82	392.54

Lately, Koutsoyiannis et al. (2008) indicated the use of analogue modeling techniques for flow prediction which give impressive performance due to advances in nonlinear dynamical systems (chaotic systems). The major drawbacks of these approaches are that they are data intensive and work as black boxes, thus provide no insight into the hydrological processes.

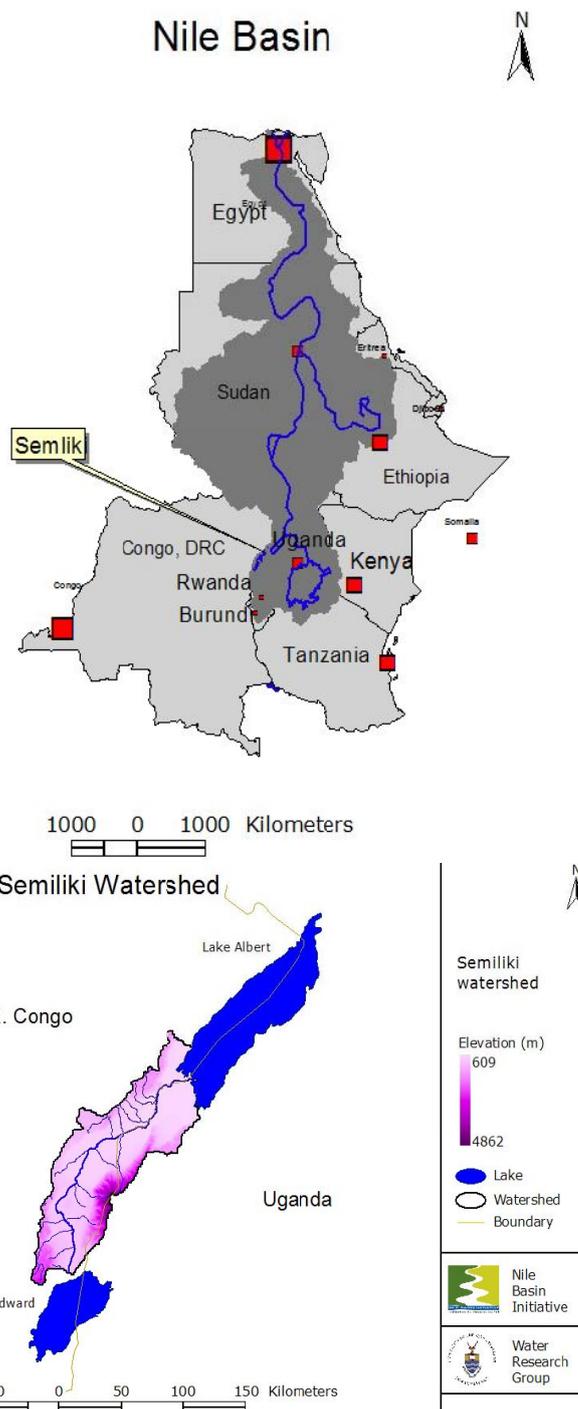
This paper reports on linear and nonlinear regression modeling approaches for flow prediction in a medium size watershed of the equatorial Nile region (Semliki catchment), where very little hydro-meteorological data are available. These approaches attempt to provide monthly flow estimates that relate similar catchment physiographic attributes to generated flows.

2 Study area

These analyses are conducted within the Semliki watershed of the equatorial Nile region (Fig. 1). The catchment studied covers an estimated area of 7699 km². The Semliki drains the basins of lakes Edwards and George, and a contributing area downstream that includes the western slopes of the Ruwenzori range. The watershed receives an average rainfall of 1245 mm per annum, with peaks occurring in May (95 mm/month) and October (205 mm/month). An average annual local runoff of 4.622 km³ has been estimated from records at Bweramule (Sutcliffe and Parks, 1999). The elevations comprise flat areas and ice-capped mountains, climbing up to 4862 m above the sea level. The flora and the fauna of the watershed constitute one of the unique and distinct ecosystems of the Albertine Rift region. The vegetation predominantly comprises medium altitude moist evergreen to semi deciduous forest. Five distinct vegetation zones have been documented under the mount Ruwenzori and they occur with changes in altitude. Detailed information on landscape physiographic attributes are reported in Kileshye Onema and Taigbenu (2009).

3 Data and methods

The landscape of any catchment is made up of several combinations of physiographic attributes. These combinations are

**Fig. 1.** Semliki watershed.

usually variable among catchments, giving rise to different hydrological responses. The Semliki catchment was delineated into 21 sub-catchments (S3–S23) that are presented in Fig. 2. This delineation was achieved using the 90m Digital Elevation Model (DEM) from the Shuttle Radar Topography Mission (SRTM) along with traditional maps. From six physiographic attributes, the 21 sub-catchments were grouped

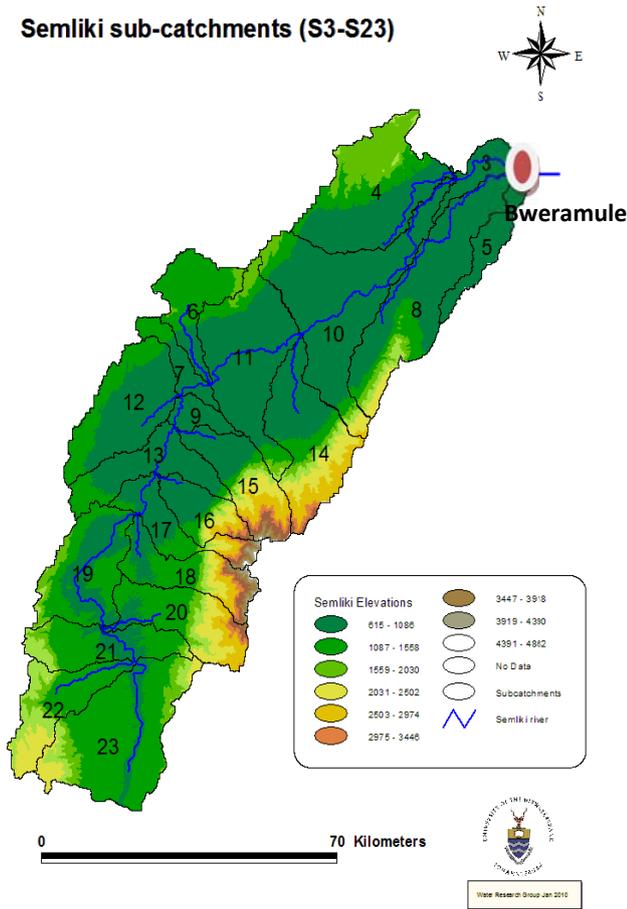


Fig. 2. Semliki Subcatchments (S3 to S23).

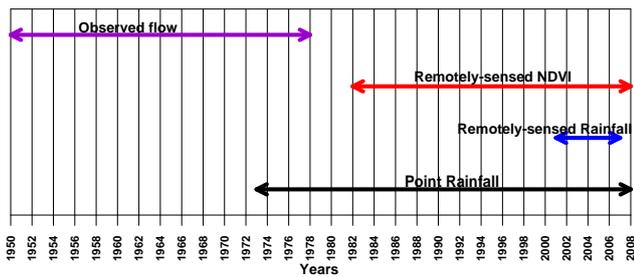


Fig. 3. Periods of available causal climate-related data (rainfall and NDVI) and flow.

into two categories using the principal component analysis (PCA) and cluster analysis. Making use of the assumption that the physiographic variables are stationary, a quasi-temporal prediction of the flow was achieved with two additional variables (NDVI and rainfall) that have both spatial and temporal variation. These two latter variables were generated for the catchment from remote-sensed data. The geographical characteristics of these sub-catchments were identified by six variables, namely is the stream length (x_1), the drainage density (x_2), the mean stream slope (x_3), the

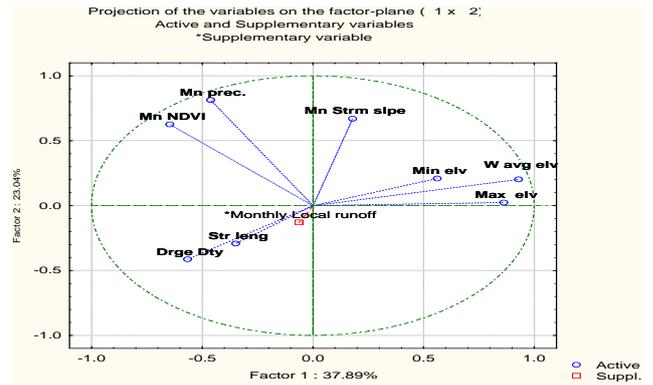


Fig. 4. PCA of variables.

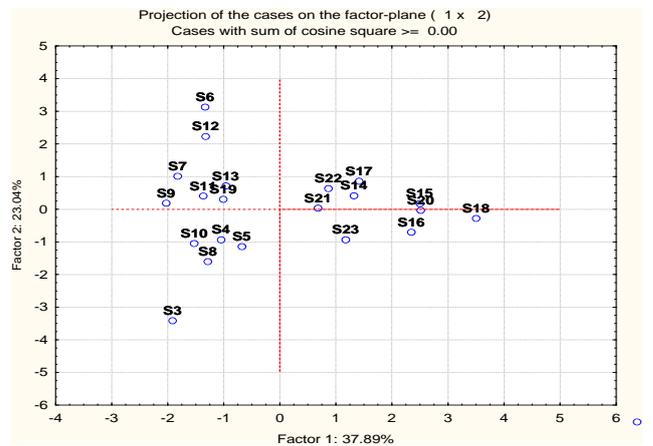
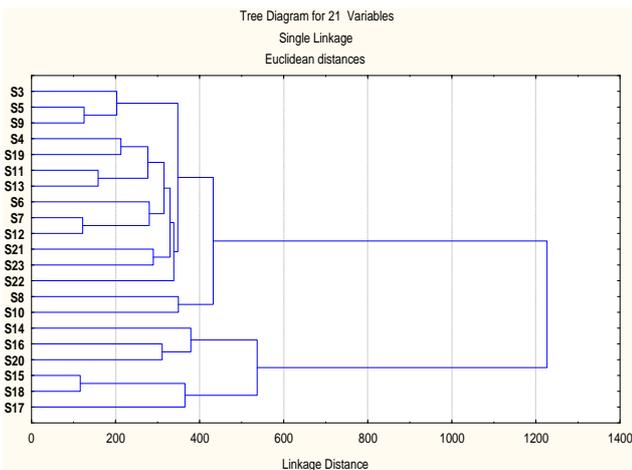


Fig. 5. Projection of cases from the PCA of variables.

maximum elevation (x_4), the minimum elevation (x_5), and the weighted average elevation (x_6). The statistical properties of these physiographic variables are presented in Table 1 and used subsequently in the grouping of these sub-catchments. In predicting monthly flows in these sub-catchments, it is assumed that these variables are stationary, constant over time but not over space. Two additional variables, generated for these sub-catchments, are the monthly rainfall (x_7) and NDVI (x_8) which have both spatial and temporal variation. The remotely-sensed NDVI data covered the period from 1982 to 2008 and were extracted from decadal maximum composite imagery provided by the National Oceanic and Atmospheric Administration-Advance Very High Resolution Radiometer (NOAA-AVHRR) and processed with the image display and analysis software WinDisp 5.1. The satellite derived rainfall data were provided by the National Oceanic and Atmospheric Administration (NOAA) through the Famine Early Warning System Network (FEWS-Net) for the period 2001–2007, and were refined by the only point daily rainfall data in the catchment obtained from a station located at Beni which covered the period from 1973 to 2008. Additional details on the processing of remotely sensed acquired datasets

Table 2. Generated landscape attributes for subcatchments (S3–S23).

Description	Source	Observation
Landform	SOTERCAF	GIS processing (WINDISP, ARC View3.3, Excel 07-10, Statistica 8.0)
Lithology	SOTERCAF	GIS processing (WINDISP, ARC View3.3, Excel 07-10, Statistica 8.0)
Soils	SOTERCAF	GIS processing (WINDISP, ARC View 3.3, Excel 07-10, Statistica 8.0)
Drainage Density (Dd)	SRTM 90m-DEM, SWAT pre-processor	Subcatchments areas generated from SWAT preprocessor (WINDISP, ARC View3.3, Excel 07-10, Statistica 8.0, SWAT)
Stream Length	SRTRM 90m-DEM, topographical map (1/50,000), SWAT pre-processor	Generated from the SWAT preprocessor and cross validation with traditional map (WINDISP, ARC View3.3, Excel 07-10, Statistica 8.0, SWAT)
Stream slope	SRTM 90m DEM	Remotely sensed acquired (ARC View 3.3, Excel 07-10, Statistica 8.0)
Rainfall	FEWS NOAA / RFE (2001–2007) rain gauge at Beni (1973 to 2008)	Remotely sensed acquired and Locally corrected and calibrated (WINDISP, ARC view3.3, Excel 07-10, Statistica 8.0)
Elevations: Minimum Maximum Area-weighted average	SRTM 90m-DEM, topographical map (1/50 000)	Remotely sensed acquired and cross validation with traditional map. (Arc View 3.3, Excel 07-10, Statistica 8.0)
NDVI	NOAA-AVHRR (1982–2008)	Remotely sensed acquired and correlated with rainfall. (WINDISP, ARC View3.3, excel 07-10, Statistica 8.0)

**Fig. 6.** Tree cluster of the physiographic variables.

can be found in Kileshye Onema and Taigbenu (2009). At the outlet of the Semliki watershed is a gauging station located at Bweramule (Fig. 2) which provided historical monthly flows from 1950 to 1978 that were used for the calibration of both

the linear and the non-linear flow prediction models that are subsequently described in this paper. The periods of available data on climate-related variables, namely rainfall and NDVI, that can be considered causal to the observed stream flow are presented in Fig. 3. The figure reveals the challenge posed in the flow modelling of this catchment which, apart from the lack of data for flows in the tributaries of the Semliki river and climatic variables for the 21 sub-catchments, the only available point rainfall data from Beni, which is not representative of the entire catchment, overlaps the record of monthly flow measurements at Bweramule for only six years from 1973 to 1978. Our problem statement, therefore, precludes the use of regionalization techniques that are based on flow duration curves (FDCs) that can be constructed by statistical or parametric or graphical approaches (Castellarin et al., 2004; Quipo et al., 1983). For each subcatchment, a historical 28-yr monthly mean volume was computed proportionally to the subcatchment area and was labeled as “control”. This historical approach is similar to the one undertaken by Asadullah et al. (2008) in data-scarce regions of central and Eastern Africa. The runoff generation approach is supported by the fact that in humid basins like the Semliki as opposed

Table 3. Coefficients of correlations between variables.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	1	0.344	-0.2	-0.05	-0.29	-0.29	-0.08	0.07
x_2		1	-0.19	-0.46	-0.24	<i>-0.45</i>	0.07	0.16
x_3			1	0.37	0.13	<i>0.54</i>	0.08	-0.15
x_4				1	0.23	<i>0.87</i>	<i>-0.5</i>	<i>-0.63</i>
x_5					1	<i>0.46</i>	-0.25	-0.23
x_6						1	-0.45	<i>-0.64</i>
x_7							1	<i>0.78</i>
x_8								1

stream length = x_1 , drainage density = x_2 , mean stream slope = x_3 , maximum elevation = x_4 , minimum elevation = x_5 , weighted average elevation = x_6 , monthly rainfall = x_7 and NDVI = x_8 . Italicized correlation coefficients are significant at 95% confidence level.

to arid and semi-arid regions, “stream flows increase in the downstream direction, and the spatial distribution of average monthly or seasonal rainfall is more or less the same from one part of the river basin to another, hence the runoff per unit land area is assumed constant over space. In these situations, estimated flows are usually based on the watershed areas, as contributing flow to those sites, and the corresponding streamflows and watershed areas above the nearest or most representative gauge sites” (Loucks et al., 1981; Loucks and Van Beek, 2005). Furthermore, rainfall for the equatorial region, one of the main driving force behind runoff generation has not changed in a statistical significant way since 1950 despite reported seasonal and interannual variations (Nicholson and Entekhabi, 1987; Bigot et al., 1998; Lienou et al., 2008). Table 2 provides landscape attributes, their sources and software used and guides any further work that intends to follow the approach undertaken in this paper.

The data generated from the physiographic variables of the 21 sub-catchments were examined to explore their groupings and similarities in their data structure. To achieve these objectives, the principal component analysis (PCA) and cluster analysis were used as exploratory techniques to study the structure and similarities of the data. The PCA analysis performed in this study used the scale invariant correlation matrix of the variables as opposed to the covariance matrix. The scaled variables are obtained by the expression:

$$X_i = \frac{x_i - \mu_i}{\sigma_i} \tag{1}$$

where X_i is the scaled variable of the original variable x_i , and μ_i and σ_i are the mean and standard deviation of the original variable.

The Glaeson-Staelin redundancy statistic and the Bartlett’s sphericity test were performed as pre-requisite to the PCA. The Glaeson-Staelin redundancy statistic (ϕ) measures the level of interrelation between a group of variables (Magingxaa et al., 2009) with a zero value indicating uncorrelated variables and unity indicating perfect correlation between the variables (Jolliffe, 2002). The Bartlett’s sphericity test is used to test the null hypothesis that the variables of the correlation matrix are not correlated (Sousa et al., 2007). When p-values

Table 4. Eigenvalues of components.

No.	Individual Eigenvalue	Cumulative Percent	Percent
1	3.49	43.59	43.59
2	1.56	19.46	63.04
3	1.02	12.71	75.75
4	0.78	9.80	85.55
5	0.64	7.97	93.51
6	0.30	3.74	97.25
7	0.17	2.13	99.38
8	0.05	0.62	100.00

Table 5. Eigenvectors of principal components.

Variables	Factor1	Factor2	Factor3
x_1	-0.16	-0.52	-0.42
x_2	-0.28	-0.36	-0.10
x_3	0.24	0.37	-0.60
x_4	0.47	-0.08	-0.27
x_5	0.27	0.17	0.56
x_6	0.51	0.10	-0.14
x_7	-0.34	0.52	-0.22
x_8	-0.42	0.37	-0.05

of the Bartlett’s test are greater than 0.05 it is not prudent to proceed with the PCA (Peres-Neto et al., 2005). The Kaiser criterion was used to identify the number of principal components for consideration as providing significant variance in the data (Jolliffe, 2002). The general suggested rule associated with this criterion is to retain principal components with eigenvalue greater than 1 (Chen and Chen, 2003), and to consider those less than 1 as trivial.

The monthly local runoff of the catchment was predicted by linear and nonlinear regression models that make use of the physiographic and meteorological variables. They are described by the relationships below. For the linear model

$$Q = a_0 + A_i^T x_i \tag{2}$$

where Q represents the monthly flows, the superscript T denotes the transpose, a_0 and $A_i^T = \{\alpha_1, \alpha_2, \dots, \alpha_8\}$ are the regression parameters, and $x_i^T = \{x_1, x_2, \dots, x_8\}$. The expression for the flow by the nonlinear model is given by

$$Q = b_0 + B_{ij}^T z_{ij} \tag{3}$$

where

$$B_{ij}^T = \{ \beta_{11}, \beta_{21}, \dots, \beta_{81} \} \text{ and } z_{ij}^T = \{ x_1, x_2, \dots, x_8 \} \tag{4}$$

To evaluate the performance of these two models, the Nash-Sutcliffe efficiency (NSE), the percent bias (PBIAS) and the

Table 6. Grouping of Semliki sub-catchments from the PCA.

Group I	Group II
S3	S14
S4	S15
S5	S16
S6	S17
S7	S18
S8	S20
S9	S21
S10	S22
S11	S23
S12	
S13	
S19	

RMSE to observation standard deviation ratio (RSR) are used. NSE is a normalized index that defines the relative magnitude of the residual variance compared to that of the measured, and expressed as (Nash and Sutcliffe, 1970)

$$NSE = 1 - \frac{\sum_{i=1}^N (Q_i^{obs} - Q_i^{sim})^2}{\sum_{i=1}^N (Q_i^{obs} - Q_i^{mean})^2} \quad (5)$$

where Q_i^{obs} and Q_i^{mean} are respectively the i th observed flow and its mean, Q_i^{sim} is the i th simulated flow, and N is the number of observations. NSE has a range between $-\infty$ and 1, with its optimal value of 1. Negative values of NSE indicate poor performance of the model, suggesting that the observed mean is a better predictor of the flows than those of the model (Moriasi et al., 2007). The PBIAS is expressed as

$$PBIAS = \frac{\sum_{i=1}^N (Q_i^{obs} - Q_i^{sim})}{\sum_{i=1}^N Q_i^{obs}} \times 100 \quad (6)$$

The optimal value of PBIAS is 0, with low-magnitude values indicating good predictive ability of the model. Positive values indicate underestimation of the observed flows by the model, and conversely for negative values (Gupta et al., 1999). The RMSE-observation standard deviation ratio (RSR) is an error statistics that normalizes the RMSE with the standard deviation of the observed flows. It is expressed as

$$RSR = \frac{RMSE}{STDEV_{obs}} = \frac{\sqrt{\sum_{i=1}^N (Q_i^{obs} - Q_i^{sim})^2}}{\sqrt{\sum_{i=1}^N (Q_i^{obs} - Q_i^{mean})^2}} \quad (7)$$

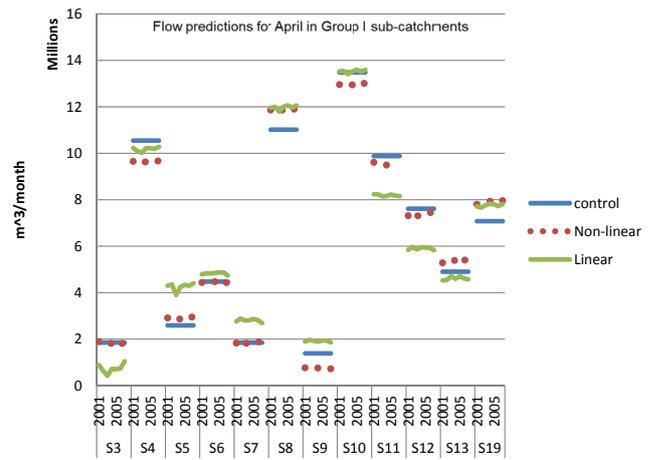


Fig. 7. Predicted flows for April in Group I sub-catchments.

The RSR is a normalized error index that utilizes the benefits of the RMSE which is one of the frequently used error index statistics. Excellent model performance will produce a zero value of RSR or RMSE (Moriasi et al., 2007).

4 Results and discussions

4.1 Principal Components Analysis (PCA)

The scaled variables of the physiographic and meteorological variables, expressed by Eq. (1), are incorporated into the PCA. In this way the correlation matrix of the variables is scale invariant. The correlations between the variables are summarized in Table 3. Italicized correlation coefficients are significant at 95 % confidence level. There are some high correlations (greater than 0.5), implying that there is a correlation structure that can potentially be modeled or further explored using the PCA. The Glaeson-Staelin redundancy test yields a value of $\phi = 0.395$, suggesting that there is considerable complexity in the data of these variables which warrants further examination using the PCA. The p-value from Bartlett’s sphericity test is 0.00000 which indicates that the null hypothesis can be rejected and there is significant strength in the relationship among the variables to warrant carrying a PCA.

The eigenvalues of the principal components are presented in Table 4, and making use of the Kaiser criterion allows for the retention of the first three principal components which account for 76 % of the variation in the data. The factor loadings of these three principal components reflect the contributions and roles of these variables in the correlation of the data. The eigenvectors of the three loading factors are presented in Table 5. The factor loadings are the correlations between the variables and the factors. Factor 1 showed best correlation with the maximum elevation and the average elevation, whereas factor 3 showed best correlation with the

Table 7. Dimensionless Performance statistics (NSE, PBIAS, RSR).

		Linear model			Nonlinear model		
		NSE	PBIAS	RSR	NSE	PBIAS	RSR
Group I (Flat)	Jan	0.933	0.097	0.259	0.977	-3.816	0.152
	Feb	0.932	0.032	0.260	0.982	0.608	0.132
	Mar	0.930	-4.449	0.265	0.982	0.125	0.136
	Apr	0.932	-0.030	0.260	0.982	0.156	0.135
	May	0.932	0.023	0.261	0.981	0.623	0.137
	Jun	0.934	0.023	0.257	0.982	-0.178	0.134
	Jul	0.934	0.020	0.256	0.982	0.468	0.133
	Aug	0.933	-0.004	0.258	0.982	0.059	0.135
	Sep	0.937	0.004	0.251	0.982	-0.483	0.133
	Oct	0.934	0.025	0.257	0.982	0.461	0.133
	Nov	0.938	0.027	0.249	0.982	0.276	0.135
	Dec	0.934	0.016	0.256	0.982	-0.199	0.135
Group II (Hilly)	Jan	0.997	-1.060	0.057	0.996	-0.236	0.061
	Feb	0.998	0.006	0.049	0.997	0.198	0.057
	Mar	0.998	0.009	0.049	0.997	0.071	0.057
	Apr	0.998	-0.015	0.050	0.997	0.054	0.057
	May	0.997	-0.905	0.054	0.996	1.192	0.066
	Jun	0.991	-3.094	0.097	0.997	0.173	0.057
	Jul	0.998	0.008	0.050	0.997	-0.232	0.055
	Aug	0.998	-0.042	0.049	0.997	0.232	0.058
	Sep	0.998	-0.003	0.049	0.997	0.375	0.059
	Oct	0.998	0.001	0.049	0.997	-0.119	0.058
	Nov	0.996	1.638	0.063	0.997	-0.049	0.059
	Dec	0.998	-0.006	0.049	0.997	-0.351	0.059

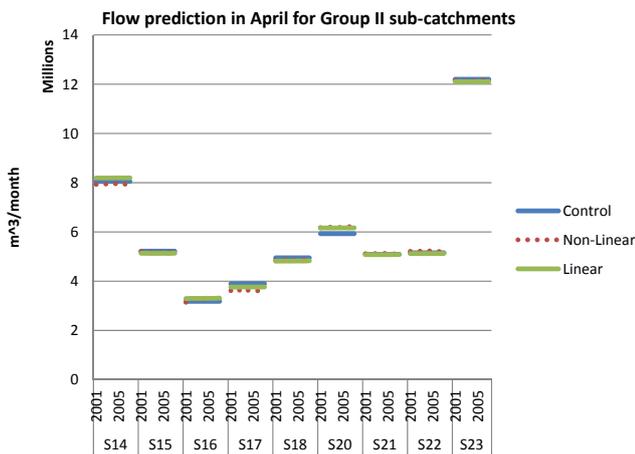


Fig. 8. Predicted flows for April in Group II sub-catchments.

average slope and the minimum elevation. Figure 4 provides an illustration of the projection of the variables on a factor plane using an alternative criteria for the PCA of the variables. Each quadrant represents a similar group of variables. Further use of the PCA is made in order to identify groupings of sub-catchments by assessing the projection of cases onto the factor plane (Fig. 5). The figure establishes four groups of

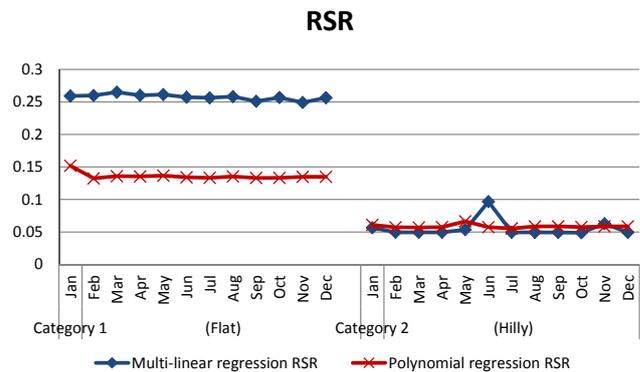


Fig. 9. RSR Performance statistics.

sub-catchments that are further reduced to two main groups or categories (Table 6) so as to simplify the prediction equations of the runoff, which is the main goal of this paper. The physiographic attributes, namely mean stream slope, minimum elevation, maximum elevation, weighted average elevation that provide this major categorization from the PCA are located in the first quadrant of Fig. 4. All sub-catchments in group I, except S19, are characterized by low elevations, while those in group II are generally characterized with high elevations with weighted values that are above 1122 m. The

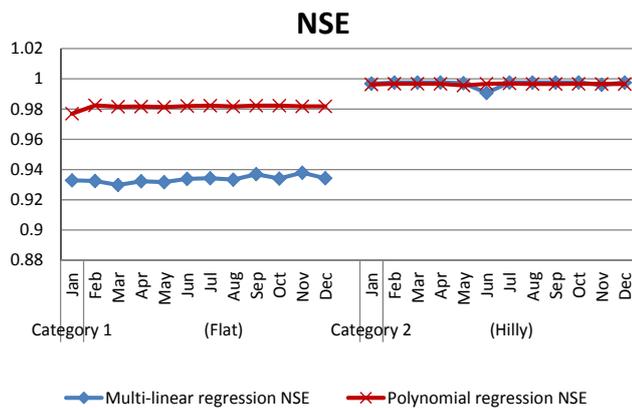


Fig. 10. NSE Performance statistics.

results of the tree cluster analysis provide a similar trend as that of the PCA, except for the discrepancy in the grouping of S21, S22 and S23 (Fig. 6).

4.2 Linear and non-linear model simulation results

As earlier indicated, the six physiographic variables are assumed stationary while the two meteorological variables have spatial and temporal (at monthly time step) variations. The regression parameters of the models are evaluated using Eqs. (2) and (3) from which estimates of the monthly local runoff are obtained. The predicted flows from the two models are of order of magnitude comparable to estimates provided by Senay et al. (2009) in their attempt to document the overall basin dynamics of the Nile River. Typical mean monthly flow hydrographs for the month of April (considered a wet month) are presented in Fig. 7 for group I sub-catchments from 2001–2007 and Fig. 8 for group II sub-catchments. “Control” values for flows were computed as reported under Sect. 3 of this paper. The values of the performance error statistics (NSE, PBIAS and RSR) presented in Table 7 for the predicted monthly flows in the sub-catchments indicate very satisfactory performance by both models, though the nonlinear model has a better performance than the linear one especially in group I of subcatchments (Figs. 9 and 10).

5 Conclusions

This study reports on the use of two modelling approaches for the prediction of the monthly flows in the data-scarce Semliki watershed of the equatorial Nile. The principal component analysis was carried to identify variables that explained most of the variance in the dataset which comprised six physiographic and two climate-related variables. Similar sub-catchments were grouped into two categories on the basis in of their physiographic attributes, and monthly runoff was then generated by the linear and nonlinear regression models. The dimensionless statistics (NSE, PBIAS and

RSR) for the two models indicate that the nonlinear model outperforms the linear one especially in the first group of subcatchments characterised by flatter elevations. Whereas these two mesoscale and deterministic models do not in any way directly address the hydrological processes in the catchment, they provide a statistical relationship between landscape attributes and monthly flows. The order of magnitude of these estimates that relate similar physiographic attributes to monthly flows can be useful for preliminary assessment of the water resources of a very poorly gauged catchment like the Semliki. The results from the current work could subsequently be complemented by a physically-based hydrological approach that explicitly accounts for the interaction between landscape attributes and water resources in the Semliki watershed, but that would still require good hydro-meteorological data.

Acknowledgements. The support from the Applied Training Project of the Nile Basin Initiative (ATP/NBI) and WaterNet is highly appreciated. The authors extend also their gratitude to the reviewers of the initial version of this manuscript. The opinions and results presented in this paper are those of the authors and do not necessarily represent the donors or participating institutions.

Edited by: A. Castellarin

References

- Asadullah, A., McIntyre, N., and Kigobe, M.: Evaluation of five satellite products for estimation of rainfall over Uganda/Evaluation de cinq produits satellitaires pour l'estimation des précipitations en Ouganda, *Hydrolog. Sci. J.*, 53, 1137–1150, 2008.
- Bigot, S., Moron, V., Melice, J. L., Servat, E., and Paturel, J. E.: Fluctuations pluviométriques et analyse fréquentielle de la pluviosité en Afrique centrale, in: *Water Ressources Variability in Africa during the XXth Century (Proc. Abidjan Symp., November 1998)*, edited by: Servat, E., Hughes, D., Fritsch, J.-M., and Hulme, M., 215–222, IAHS Publ. 252, IAHS Press, Wallingford, UK, 1998.
- Castellarin, A., Galeati, G., Brandimarte, L., Montanari, A., and Brath, A.: Regional flow-duration curves: reliability for ungauged basins, *Adv. Water Resour.*, 27, 953–965, 2004.
- Chen, D. and Chen, Y.: Association between winter temperature in China and upper air circulation over East Asia revealed by canonical correlation analysis, *Global Planet. Change*, 37, 315–32, 2003.
- Cheng, Q., Ko, C., Yuan, Y., Ge, Y., and Zhang, S.: GIS modeling for predicting river runoff volume in ungauged drainages in the greater Toronto area, Canada, *Comput. Geosci.*, 32, 1108–1119, 2006.
- Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration, *J. Hydrol. Eng.*, 4, 135–143, 1999.
- Heuvelmans, G., Muys, B., and Feyen, J.: Regionalization of the parameters of a hydrological model: comparison of linear regression models with artificial nets, *J. Hydrol.*, 319, 245–265, 2006.

- Jolliffe, I. T.: Principal Component Analysis, Springer series in statistics, ISBN 0-387-95442-2, New York, USA, 2002.
- Kileshye Onema, J.-M. and Taigbenu, A. E.: NDVI–rainfall relationship in the Semliki watershed of the equatorial Nile, *Phys. Chem. Earth*, 34, 711–721, 2009.
- Koutsoyiannis, D.: Uncertainty, entropy, scaling and hydrological stochasticity. 1 Marginal distributional properties of hydrological processes and state scaling, *Hydrolog. Sci. J.*, 50, 381–404, 2005a.
- Koutsoyiannis, D.: Uncertainty, entropy, scaling and hydrological stochasticity. 2 Time dependence of hydrological processes and time scaling, *Hydrolog. Sci. J.*, 50, 405–426, 2005b.
- Koutsoyiannis, D., Yao, H., and Georgakakos, A.: Medium-range flow prediction for the Nile: a comparison of stochastic and deterministic methods, *Hydrolog. Sci. J.*, 53, 405–426, 2008.
- Kwon, H.-H., Brown, C., Xu, K., and Lall, U.: seasonal and annual maximum streamflow forecasting using climate information: application to the Three Gorges dam in the Yangtze river basin, China, *Hydrolog. Sci. J.*, 54, 606–622, 2009.
- Lienou, G., Mahé, G., Paturel, J., Servat, E., Sighomnou, D., Ekodeck, G. E., Dezetter, A., and Dieulin, C.: Evolution des régimes hydrologiques en région équatoriale camerounaise: un impact de la variabilité climatique en Afrique équatoriale?, *Hydrolog. Sci. J.*, 53, 789–801, 2008.
- Loucks, D. P. and Van Beek, E.: Water Resources Systems Planning and Management, UNESCO publications, ISBN 92-3-103998-9, Delft Hydraulics, The Netherlands, 2005.
- Loucks, D. P., Stedinger, J. R., and Haith, D. A.: Water resource systems planning and analysis, Prentice-Hall, Inc, Englewood Cliffs, New Jersey, 07632, 1981.
- Magingxaa, L. L., Alemub, Z. G., and Van Schalkwyk, H. D.: Factors influencing access to produce markets for smallholder irrigators in South Africa, *Development Southern Africa*, 26, 47–58, doi:10.1080/03768350802640081, 2009.
- McIntyre, N., Lee, H., Wheeler, H., Young, A., and Wagener, T.: Ensemble predictions of runoff in ungauged catchments, *Water Resour. Res.*, 41, W12434, doi:10.1029/2005WR004289, 2005.
- Merz, R. and Blöschl, G.: regionalization of catchment model parameters, *J. Hydrol.*, 287, 95–123, 2004.
- Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Hardmel, R. D., and Veith, T. L.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, *American Society of Agricultural and Biological Engineers*, 50, 885–900, 2007.
- Mwakalila, S.: Estimation of stream flows of ungauged catchments for river basin management, *Phys. Chem. Earth*, 28, 935–942, 2003.
- Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models: Part I. A discussion of principles, *J. Hydrol.*, 10, 282–290, 1970.
- Nicholson, S. E. and Entekhabi, D.: Rainfall variability in equatorial and Southern Africa: Relationship with Sea Surface Temperatures along the Southwestern Coast of Africa, *J. Clim. Appl. Meteorol.*, 26, 561–578, 1987.
- Peres-Neto, P. R., Jackson, D. A., and Somers, K. M.: How many principal components? stopping rules for determining the number of non-trivial axes revisited, *Comput. Stat. Data An.*, 49, 974–997, 2005.
- Quimpo, R. G., Alejandrino, A. A., and McNally, T. A.: Regionalised flow duration curves for Philippines, *J. Water Res. Pl. ASCE*, 109, 320–330, 1983.
- Sanborn, S. C. and Bledsoe, B. P.: Predicting streamflow regime metrics for ungauged streams in Colorado, Washington and Oregon, *J. Hydrol.*, 325, 241–261, 2006.
- Schröder, B.: Pattern, process, and function in landscape ecology and catchment hydrology – how can quantitative landscape ecology support predictions in ungauged basins?, *Hydrol. Earth Syst. Sci.*, 10, 967–979, doi:10.5194/hess-10-967-2006, 2006.
- Sefton, C. E. M. and Howard, S. M.: Relationship between dynamic response characteristics and physical descriptors of catchments in England and Wales, *J. Hydrol.*, 211, 1–16, 1998.
- Senay, G. B., Asante, K., and Artan, G.: Water balance dynamics in the Nile Basin, *Hydrol. Process.*, 23, 3675–3681, 2009.
- Shao, Q., Zhang, L., Chen, Y. D., and Singh, V. P.: A new method for modeling flow duration curves and predicting streamflow regimes under altered land-use conditions, *Hydrolog. Sci. J.*, 54, 582–595, 2009.
- Sharda, V. N., Prasher, S. O., Patel, R. M., Ojasvi, P. R., and Prakash, C.: Performance of Multivariate Adaptive Regression Splines (MARS) in predicting runoff in mid-Himalayan micro-watersheds with limited data, *Hydrolog. Sci. J.*, 53, 1165–1175, 2008.
- Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J. J., Mendiondo, E. M., O'Connell, P. E., Oki, T., Pomeroy, J. W., Schertzer, D., Uhlenbrook, S., and Zehe, E.: IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences, *Hydrolog. Sci. J.*, 48, 857–880, 2003.
- Sousa, S. I. V., Martins, F. G., Alvim-Ferraz, M. C. M., and Pereira, M. C.: Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations, *Environ. Model. Softw.*, 22, 97–103, 2007.
- Sutcliffe, J. V. and Parks, Y. P.: The hydrology of the Nile, IAHS special publication No. 5, 1999.
- Uhlenbrook, S. and Siebert, A.: On the value of experimental data to reduce the prediction uncertainty of process-oriented catchment model, *Environ. Model. Softw.*, 20, 19–32, 2005.
- Xu, C.: Testing the transferability of regression equations derived from small sub-catchments to a large area in central Sweden, *Hydrol. Earth Syst. Sci.*, 7, 317–324, doi:10.5194/hess-7-317-2003, 2003.
- Yadav, M., Wagener, T., and Gupta, H.: Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, *Adv. Water Resour.*, 30, 1756–1774, 2007.
- Zhang, Z., Wagener, T., Reed, P., and Bhushan, R.: Reducing uncertainty in prediction in ungauged basins by combining hydrologic indices regionalization and multiobjective optimization, *Water Resour. Res.*, 44, W00B04, doi:10.1029/2008WR006833, 2008.