1 Introduction

Some of the formidable challenges that the Nile Basin faces include: floods and droughts due to climate variability, restive trans-boundary management issues, widespread poverty, high demographic growth rates, food insecurity resulting from a combined effect of rainfall variability and the unswerving dependence of majority of the population on subsistence and rain-fed agriculture, etc. To deal with these challenges, sufficient planning and management of the River Nile water resources is required. One form of support to such requirement is the enhanced comprehension of the historical patterns of flow variation and their spatial differences across the entire Nile Basin as done in this study. So far several investigations were made on the variability of flow and hydro-climatic variables for the Nile Basin.

Although the variation of river flows in the Nile Basin may be ascribed to the changes in rainfall, a number of studies based on remotely sensed land cover or satellite data (see e.g. Elmqvist, 2005; Rientjes et al., 2011) or aerial photographs (Bewket and Sterk, 2005) have reported on the effects of anthropogenic factors on river flow regimes. According to Rientjes et al. (2011), forest cover decreased from 50 to 16% in the Gilgel Abay catchment in the Lake Tana basin over the period 1973–2001. Elmqvist (2005) noted that the cropland per household reduced from 0.4 to 0.1 km^2 over the period 1969–2002 in Sararya Makawi, Sudan. Bewket and Sterk (2005) concluded on an increase in cultivation area in the Chemoga catchment for the period between 1960 and 1999. Flow changes were in these studies attributed to the land use changes. Bewket and Sterk (2005), for instance, related the identified decrease of 0.6 mm/year in the Chemoga catchment flows during the dry season (October to May) between 1960 and 1999 to the increase in cultivation area. The main problem with such flow change attribution studies is that for an accurate analysis, archives are required of aerial photos or satellite images of land cover with high spatial and temporal resolutions and with good quality for long time periods. Such archives are difficult to obtain for the study area. To partly meet the limitation of such archives, some studies complemented the available land use and cover data from satellite images with catchment hydrological modelling. The effect of the change in catchment characteristics on the watershed hydrology can indeed be investigated using hydrological models, by preference to fully distributed process-based models. However, the input data required by such detailed hydrological models are of large amount. Besides, due to their structural complexity and over-parameterization, the parameters of such models are difficult to optimally estimate. Alternatively, conceptual models that are more parsimonious hence with fewer parameters than the physically-based models can be applied to assess changes in catchment response in a meteorological-river flow data-based way, but at a lumped catchment scale. Such modelling studies were conducted by Mango et al. (2011), and Olang and Fürst (2011) for the equatorial region; Legesse et al. (2003, 2004), Bewket and Sterk (2005), Rientjets et al. (2011), and Gebrehiwot et al. (2013) for Ethiopia. Based on the land-use scenario investigation using the Soil and Water Assessment Tool, Mango et al. (2011)

concluded for the Mara catchment that the magnitude of the extreme low/high flows would reduce/increase if the conversion of forests to agriculture and grassland in the headwaters of the catchment continued. Olang and Fürst (2011) used the HEC-HMS rainfall-runoff model to investigate the effect of the land-use changes over the period between 1973 and 2000 on the hydrology of the Nyando catchment. The authors found an increase of 16% in the peak discharges over the entire period considered. Using the PRMS model, Legesse et al. (2003) found that flow would reduce to about 8% if the dominantly cultivated/grazing land of South Central Ethiopia was to be converted to woodland. Similarly for Lake Abiyata, Legesse et al. (2004) noted a remarkable mismatch between the observed and PRMS-modeled lake level over the period 1984–1996 compared with that for 1968–1983. The authors ascribed this discrepancy to human influence on the lake in terms of the direct use of the influent rivers. By dividing the time series over the period 1960–2004 into three parts based on either the political and land management policy changes, Gebrehiwot et al. (2013) applied the HBV model to investigate the effect of land-use changes on the runoff flows in the Birr, Upper-Didesa, Gilgel Abbay, and Koga catchments of the Blue Nile Basin. According to the authors, although six out of nine parameters of the HBV model changed significantly over the three periods during the rainfall-runoff modeling, the integrated functioning of the watersheds showed minimal changes.

The problem with the above studies each of which applied only one hydrological or rainfall-runoff model lies in the lack of insight about the influence of the model selection on the conclusive flow variation attribution. Based on the model complexity and set of parameters for calibration, the judgment of the confidence in the selection of a particular model to investigate the effect of land-use change on the flow variation is not a simple task. Moreover, other factors such as the change in meteorological conditions need to be addressed as well. Studies by Abtew et al. (2009), Camberlin (1997), Taye and Willems (2013), Tierney et al. (2013) gave evidence that the variability in hydro-climatic variables such as rainfall over the study area can be explained by the variations in large-scale ocean-atmosphere interactions.

In this study, because of the data limitation and quality problem for rainfall-runoff modeling in the Nile Basin, three rainfall-runoff models NAM (Danish Hydraulic Institute DHI, 2007; Madsen, 2000), HBV (Bergström, 1976; AghaKouchak and Habib, 2010; AghaKouchak et al., 2013) and VHM (Willems, 2014; Willems et al., 2014) were applied. These three models were adopted in this study because they have been recently used by Taye and Willems (2013) (for NAM and VHM), and Gebrehiwot et al. (2013) (for HBV) to successfully investigate the effect of land-use change on the flow regimes in the study area. To limit the influence of subjectivity in the model calibration process and the address the models' uncertainties, the Generalized Likelihood Uncertainty Estimation (GLUE) of Beven and Binley (1992) was adopted. The model-based findings moreover were complemented with the analysis of trends and temporal variations in the observational time series to support the hypothesis of flow variation attribution. In explaining the identified trends and temporal variations, special attention was given to the co-variation of flow and rainfall. The final goal was to provide new insights in the spatiotemporal variation of annual and seasonal flows along the main rivers, considering the entire Nile Basin as study region.

More specifically, this study aimed at: 1) analyzing the spatiotemporal variation in annual and seasonal flows along the main rivers of the Nile Basin, 2) investigating the co-variation of flow and rainfall, and 3) rainfall-runoff modeling to investigate the evidence of changes in rainfall-flow catchment response behavior.

3.4 Detection and Attribution of Changes in the Flow

According to Merz et al. (2012), the flow change attribution to assumed drives can be done quantitatively in either data-based or simulation-based way. Both ways were considered and combined in this study.

3.4.1 Data-Based Approach

A data-based approach was implemented at both regional and catchment scales by comparing the correlation between the variation of flow with that of the rainfall series. High correlation means that the influence of the anthropogenic factors on the catchment runoff generation processes are limited.

3.4.2 Simulation-Based Approach

The data-based approach was complemented with a simulation-based approach, where three models were applied to study the catchment rainfall-runoff flow co-variation considering a lumped catchment approach. In case of an unchanging catchment behavior, hence in case of insignificant anthropogenic factors, the temporal flow variations are assumed to be fully described by the variations in the meteorological model inputs (rainfall and evapotranspiration) after keeping the model parameters constant over time. On the other hand, in case of a change in catchment behavior due to anthropogenic influence, there would be a temporal change in the difference between the observed and modeled runoff flows and sub-flows when model parameters are kept constant. Anthropogenic influences such as deforestation, overgrazing, significant expansion of urbanized areas etc over a given catchment would: 1) affect the amount of infiltration into the soil, 2) alter the amount and velocity of the overland flow, 3) modify the rate and amount of evaporation, etc. Hence, these would alter the catchment response to the rainfall input. This difference in response should be visible through the changes in runoff volumes, sub-flow volumes, ratio between sub-flow volumes, model parameters describing the sub-flow response to such times such as the recession constant.

Because of the importance to study the runoff sub-flows and more specifically the overland flow separately, a numerical digital filter was applied to split the flow into the various sub-components. This discharge splitting was done based on the sub-flow recession constants as applied in the tool provided by Willems (2009). The simulation-based approach to search for the temporal changes in the overland flow was analyzed using three approaches including the Cumulative Rank Difference (CRD) (Onyutha, 2016a) technique, the Quantile Perturbation Method (QPM) (Ntegeka and Willems, 2008; Willems, 2013), and the well-known Mann-Kendall (MK) (Mann, 1945; Kendall, 1975) test. The CRD and QPM were applied directly to the annual maxima, annual minima and annual mean flow. The MK test was conducted on the model residuals. Each of the three methods CRD, QPM and MK analyzes the given data in a different way. Whereas the CRD focuses on the cumulative effects of the variation, the QPM considers quantile changes, and the MK deals with trends.

3.4.2.1 The Cumulative Rank Difference Method

Severe events tend to temporarily occur in the form of clusters above or below the long-term mean (call it the reference) of the hydro-meteorological variable. The tendency of the variable over time to cross the reference from below to above or *vice versa* brings about positive and negative effects on the trend in the full time series. Because these positive and negative effects cancel out each other within the dataset, the overall trend from the full time series is consequently the net effect of such cancellations (Onyutha et al., 2015). The identification and assessment of the sub-trends (i.e. the short-durational trend directions within the series) is vital to ascertain the possibility of any intervention of climate fluctuations on the hydro-meteorological variable (Onyutha, 2016a). Detection of changes in a purely statistical way using the full-time series might yield results which are meaningless sometimes (Kundzewicz and Robson, 2000) and moreover, it can disregard the occurrences and significance of the sub-trends (if any) within the dataset which may be of interest to an environmental practitioner. To graphically reveal the hidden short-durational changes (e.g. jumps in the mean, sub-trends, etc) within the time series, the CRD plot was used. To construct the CRD plots for observed and modeled flows, the following steps were taken:

i) rescaling of the given series in a nonparametric way using Eq. (2) to obtain the difference (D) between the exceedance and non-exceedance counts of the data points;

$$D(i) = 2R_a(i) - \left(n - w(i)\right) \qquad \text{for } 1 \le i \le n \qquad (2)$$

where R_a is the number of times a data point is exceeded, and w the number of times a data point appears within the given sample. To determine R_a or w, each data point is counted as if it was not considered before (Onyutha, 2016b). Considering the hypothetical series (3, 6, 3, 7, 9, 5, 3, 5); n = 8 and for i = 1 to n, R_a = (5, 2, 5, 1, 0, 3, 5, 3), w = (3, 1, 3, 1, 1, 2, 3, 2), and D = (5, -3, 5, -5, -7, 0, 5, 0).

ii) calculating cumulative sum (S_m) of the rank difference using Eq. (3);

$$S_m(i) = \sum_{j=1}^{i} D(j) \qquad \text{for } 1 \le i \le n \qquad (3)$$

For the series in Step (i), the $S_m(i)$ = (5, 2, 7, 2, -5, -5, 0, 0); it can be checked that $S_m(n)$ or $\sum_{j=1}^{n} D(j)$ is always zero.

iii) making the CRD plot i.e. plotting the $S_m(i)$ against the time unit of the series;

iv) identifying the short-durational changes from the CRD using the graphical guidelines following Onyutha (2016a), in Figure A1 generated based on synthetic series Y of n = 200.

In the CRD plot (see illustration in Figure A1), taking S_m = 0 line as the reference, the values above or below this reference are considered to characterize sub-trends in the series (Onyutha, 2016a). If the given series is characterized by an increasing trend in the first half and a decrease in the second half, for example, two curves are formed such that the first one (first half of the period) is above the reference and the second one below the S_m = 0 line (see case (a) and (1) of Figure A1). When there is no trend in the data, the CRD curve crosses the reference a number of times with no clear area over large time period between the curve and the S_m = 0 line (see case (b) and (2) of Figure A1). For a positive/negative trend, most if not all the scatter points in the CRD plot take the form of a curve above/below the reference (see case (c) and (3) of Figure A1). For a step upward/downward jump in the mean of the series (so long as there is no trend in both parts of the sub-series before and after the step jump), the scatter points take the form of two lines which meet at a point (call it the vertex) above/below the reference (see case (d) and (4) of Figure A1). For an upward/downward jump, the slope of the first line is positive/negative while that for the second one is negative/positive. For further details on the use of the CRD plots to identify changes in the series, the reader is referred to Onyutha (2016a).

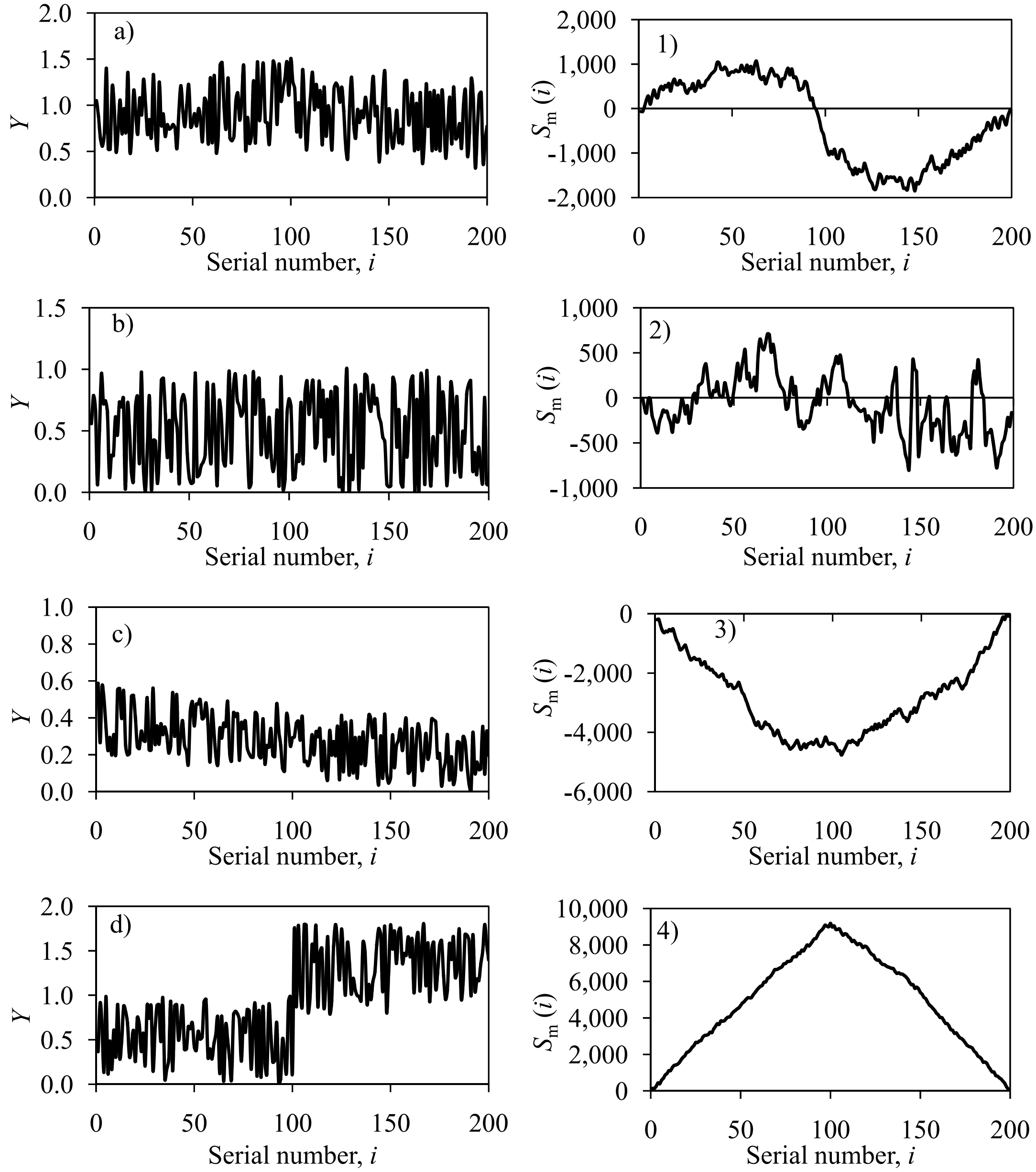

Figure A1 The CRD plots for various forms of changes in the synthetic series, Y of n = 200. The corresponding CRD plot for the series in a)-d) are shown in 1)-4) respectively.

3.4.2.2 The Quantile Perturbation method

Unlike the CRD method which relies on rescaled series, the QPM uses the given series directly (i.e. without rescaling) to obtain quantile anomalies. This allows the QPM outputs to be importantly applicable, for instance, in revising design quantiles to account for the decadal or multi-decadal oscillations or variability in the hydro-meteorological variable. To apply the QPM, two series are derived from the same data set. One series (call it y) is the full series, and the other (denoted by x) is a sub-set extracted as a sub-period from the full series. The sub-series are contained in a moving window of a specified block length (taken as 15 years in this study). The moving window is first put at the beginning of the full time series and afterwards moved by 1 year at a time. For each moving window, quantile perturbation factors are computed as the quantiles above a selected exceedance probability threshold selected from x and divided by their corresponding counterparts from y. The ultimate anomaly for the window under consideration is determined as the average of the perturbation factors for all empirical quantiles above a given threshold. The ultimate anomalies from the different moving window positions are considered to characterize the variability of the extreme quantiles in the series. An elaborate and systematic description of the QPM can be obtained from Ntegeka and Willems (2008) and Willems (2013).

3.4.2.3 The Mann-Kendall Test

For the MK test, the data points were assigned scores based on the comparison of their magnitudes. A score of +1 (or -1) was assigned if the most recent data point was larger (or smaller) than the previous one. If the two consecutive data points were equal, a score of zero was assigned. The sum of the scores for the full time series was taken as the MK test statistic S. A positive/negative value of S was taken to indicate an increasing/decreasing trend. The influence of auto-correlation on the variance of S was corrected based on the procedure suggested by Yue and Wang (2004). Next, the statistic S was standardized to obtain Z which follows the standard normal distribution with the mean of zero and variance of one. Trend in the series was considered insignificant if Z was less than the standard normal variate $Z_{\alpha/2}$ where α% is the significance level; otherwise, the trend was significant. In the simulation-based procedure, with the premise that the models fully capture the catchment behavior, any deviation between observed and simulated flow can be attributed to internal disturbances which may include forest cover change, urbanization, river engineering, dam construction, etc (Harrigan et al., 2014). In this study, any persistent deviation between the observed and modeled flow was deemed to be reflected in the occurrence of significant trend in the model residuals.

The block of text below on the MK test results will be inserted in line 19 (page 12186) of Section 4.6 in the Discussion Paper.

Table 10 shows the statistic Z values of the MK test conducted on model residuals computed based on the annual maxima, annual minima and annual mean flow. At the significance level of 5% level, the threshold $Z_{\alpha/2}$ is ± 1.96. For the Blue Nile Basin, it is noticeable that that the magnitude of Z was less than the absolute value of 1.96 for all the models. This shows that the trends in the residuals was statistically insignificant at 5% level. The trend in the residuals from the annual mean flow of Kagera was significant at 5% level for NAM. Although this might suggest changes in catchment behavior, it was not deemed conclusive since the results from the two other models i.e. HBV and VHM were insignificant. Generally, the magnitudes of Z for Kagera was greater than those of the Blue Nile. This might be due to the poor performance of the models for Kagera especially in the validation process. In some cases, the residual trend directions were also different among the models. This could be because of the difference between the models in terms of their structures and sets of parameters used to capture the runoff generation dynamics.

Table 10 Statistical results of trend in the model residuals

Catchment	Annual maxima residuals			Annual minima residuals			Annual mean residuals		
	VHM	HBV	NAM	VHM	HBV	NAM	VHM	HBV	NAM
Blue Nile	-0.32	-0.73	-0.87	-1.05	1.00	-1.89	-1.05	1.00	-1.89
Kagera	-0.75	1.56	-1.09	1.40	1.21	1.74	-1.50	1.89	**2.08**

Bold value is the statistic significant at the level of 5%.

According to Harrigan et al. (2014), despite the statistical detection of trends, rigorous attribution is required in decision making on long-term management and adaptation strategies. Additional to the call by Merz et al. (2012) for increased rigor in attribution through consistency, inconsistency and provision of confidence statement, Harrigan et al. (2014) suggested the method of Multiple Working Hypothesis (MMWHs) as a systematic examination of known drivers to explain the full signal of change. In line with the MMWHs, some of the working hypotheses (which were not investigated in this study) but deemed to potentially influence the catchment behavior (though insignificantly) across the study area and should be considered for further research in a combined way include: urbanization, forest cover transition, agricultural land-use and management change, etc. Another factor which cannot be ruled out in influencing the change detection is the questionable quality of hydro-meteorological data in the study area. Once the large data requirement for attribution become manageable in future, an interesting attempt would be to expose the interaction (if any) of the drivers of the flow changes in the various catchments of the Nile Basin.

Table 2 Daily and monthly rainfall stations for selected catchments

	Station		Location		Data period		Statistical metric		
	Paper ID	Name	Long.	Lat.	From	To	C_k [-]	C_s [-]	C_v [-]
Kagera Basin	Kag1	Mugera (Paroisse)	29.97	-3.32	1940	1990	0.98	0.73	0.78
	Kag2	Muyinga	30.35	-2.85	1940	1992	-0.17	0.44	0.72
	Kag3	Igabiro Estate	31.55	-1.82	1940	1994	0.47	0.78	0.80
	Kag4	Musenyi (Paroisse)	30.03	-2.97	1940	1994	1.34	1.06	0.83
Atbara Catchment	Atb1	Atbara	33.97	17.70	1907	1995	36.39	5.18	3.06
	Atb2	Ungwatiri	36.00	16.90	1950	1981	22.60	4.26	2.66
	Atb3	Abu-Quta	32.70	14.88	1948	1987	8.65	2.81	2.06
	Atb4	Haiya	36.37	18.33	1950	1981	33.39	5.10	2.67
	Atb5	Gedaref	35.40	14.03	1903	1996	3.07	1.78	1.50
	Atb6	Ghadambaliya	34.98	14.20	1948	1988	3.88	1.95	1.65
Blue Nile Basin	Blu1	Bahr Dar	37.41	11.60	1964	2004	0.95	1.36	1.30
	Blu2	Debremarcos	37.67	10.33	1964	2004	-0.46	0.86	1.00
	Blu3	Gonder	37.40	12.55	1964	2004	1.54	1.44	1.21
	Blu4	Addis Ababa	38.75	09.03	1964	2004	-0.06	0.95	1.02
	Blu5	Kombolcha	39.83	11.10	1964	2004	2.32	1.56	1.11
Kyoga Basin	Kyo1	Imanyiro	33.27	0.29	1950	1977	3.51	1.34	0.65
	Kyo2	Kapchorwa	34.43	1.24	1950	1995	3.07	1.15	0.69
	Kyo3	Buwabwale	34.21	0.54	1950	1977	4.73	1.61	0.69
	Kyo4	Ivukula	33.35	0.57	1950	1997	2.12	1.24	0.68
Monthly rainfall stations adopted from Onyutha and Willems (2015)	A	Kabale	29.98	-1.25	1917	1993	0.49	0.07	0.17
	B	Namasagali	32.93	1.00	1915	1978	1.89	6.12	0.19
	C	Igabiro	31.53	-1.78	1931	1982	0.82	0.80	0.24
	D	Kibondo	30.68	-3.57	1926	1978	2.47	9.50	0.29
	E	Ngudu	33.33	-2.93	1928	1971	1.61	3.88	0.33
	F	Shanwa	33.75	-3.15	1931	1985	0.68	0.28	0.24
	G	Tarime	34.37	-1.35	1933	1975	1.48	1.94	0.20
	H	Bujumbura	29.32	-3.32	1930	2004	0.27	0.09	0.18
	I	El-Da-Ein	26.10	11.38	1943	1990	0.27	0.17	-0.03
	J	El-Fasher	25.33	13.62	1917	1996	0.43	1.22	1.92
	K	El-Obeid	30.23	13.17	1902	1996	0.32	0.53	0.39
	L	En-Nahud	28.43	12.70	1911	1996	0.28	0.39	1.00
	M	Er-Rahad	30.60	12.70	1931	1984	0.31	0.80	3.21
	N	Fashashoya	32.50	13.40	1946	1988	0.32	0.04	-0.11
	O	Garcila	23.12	12.35	1943	1986	0.30	1.97	6.75
	P	Hawata	34.60	13.40	1941	1988	0.23	-0.19	1.35
	Q	Jebelein	32.78	12.57	1927	1988	0.29	-0.50	-0.40
	R	Kassala	36.40	15.47	1901	1996	0.30	-0.04	-0.18
	S	Kubbum	23.77	11.78	1943	1985	0.27	-0.01	-0.23
	T	Kutum	24.67	14.20	1929	1990	0.40	0.16	0.28
	U	Nyala	24.88	12.05	1920	1996	0.29	0.11	-0.30
	V	Renk	32.78	11.75	1906	1987	0.20	0.35	-0.09
	W	Shambat-Obs.	32.53	15.67	1913	1993	0.94	2.99	11.31
	X	Shendi	33.43	16.70	1937	1990	0.74	1.16	2.08
	Y	Talodi	31.38	10.60	1916	1987	0.23	1.12	3.57
	Z	Talodi-M-Agr.	30.50	10.60	1942	1985	0.21	0.41	-0.35
	AA	Umm-Ruwaba	31.20	12.80	1912	1989	0.32	1.93	9.26
	AB	Wau	28.02	7.70	1904	1990	0.16	0.26	-0.34
	AC	Combolcha	39.72	11.08	1952	1996	0.17	-0.84	0.69
	AD	Debremarcos	37.72	10.35	1954	1998	0.11	0.82	0.59
	AE	Gambela	34.58	8.25	1905	1993	0.22	0.00	0.25
	AF	Gore	35.55	8.17	1946	1996	0.20	1.40	2.15
	AG	Wenji	39.25	8.42	1951	1994	0.28	-0.76	1.92

Long. and Lat. stand for longitude [°] and latitude [°] respectively

Table 3 Coefficient of variation of annual flows at the various stations

St. no.	Group 1 stations	C_v [-]	St. no.	Group 2 stations	C_v[-]
1	Kyaka Ferry	0.35	10	Sennar	0.23
2	Jinja	0.42	11	Khartoum	0.22
3	Paara	0.48	12	El Diem	0.19
4	Kamdini	0.31	13	Babu	0.26
5	Kafu	0.73	14	Kilo 3	0.35
6	Aswa	0.77	15	Tamaniat	0.15
7	Panyango	0.55	16	Hud. + Hass.	0.16
8	Mongalla	0.40	17	Dongola	0.17
8	Malakal	0.40	18	Aswan Dam	0.15
9	Sennar	0.20			