Validation of Satellite Rainfall Estimate Based on Rainfall Observations for Sudan

Aboubaker. A. Osman[1], Barsi. B.I[2], Bashar. K.E[3]

[1] Nile Valley University, contact author, aboubkrahmed@hotmail.com
[2] Faculty of Engineering, University of Khartoum, d_barsi@hotmail.com
[3] UNESCO Chair in Water Resources, Omdurman Islamic University, basharke@hotmail.com

Abstract

Rainfall is the main driving force in the hydrology of arid and semi-arid regions. Where the rain gauge network is sparse, the complete and regular coverage provided by satellite Rainfall Estimate is a major advantage for hydrological and agricultural modeling.
The main objective of this paper is to assess and evaluate the reliability of the US National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC) Rainfall Estimate (RFE) 2.0 data using monthly rainfall observations at 14 stations covering most area of the Sudan during the rainy season. The monthly rainfall data of 10 years (2001 to 2010) are used in the study. Statistical indices are used to evaluate, compare and validate satellite rainfall data with respect to gauge rainfall data. Statistical indices used in this study are categorical verification statistics and continuous verification statistics.
The different comparisons of satellite CPC_RFE2.0 and observed rainfall showed significant average monthly correlation (76%). The average monthly bias showed underestimation in some stations while other station showed overestimation. The average RFE accuracy was found to be 0.92 with high probability of detection (POD, 0.95) and low false alarm ratio (FAR, 0.12). The Critical Success Index (CSI) was found to be 0.86 indicating correct diagnoses of rainfall events.
In conclusion, the above values of statistical indices goes well with the climatic zones of the gauging stations. The values of these indices also indicated that the monthly RFE over the Sudan can be considered a reliable rainfall source in the absence of gauge rainfall data.

Key words: RFE, statistical indices, satellite rainfall estimation.

1. INTRODUCTION

Rainfall is the main driving force in the hydrology of arid and semi-arid regions. It is commonly subjected to sporadic storms that vary greatly in time and space. Particularly where the rain gauge network is sparse, the complete and regular coverage provided by satellite rainfall estimate is a major advantage for hydrological and agricultural modeling.

A long-standing problem in the meteorological and hydrological studies is the sparse rain gauge network representing the spatial distribution of precipitation and its quantity on small scales. Therefore, satellite derived quantitative precipitation estimates are extremely useful for obtaining rainfall patterns and volumes.

The use of appropriate hydrological models with real-time satellite rainfall estimates can help mitigate flood damage, provide support to contingency planning, and provide warning to people threatened by floods (Bajracharya et al, 2014). Satellite precipitation estimates are widely used to measure global rainfall on near real-time and monthly time scales for climate studies, numerical weather prediction (NWP) data assimilation, now casting and flash flood warning, tropical rainfall potential and water resources monitoring. Therefore, similar to any observational data, investigating their accuracy and limitations is crucial. This is done by verifying the satellite estimates against independent data from gauges and radar (Levizzani et al, 2007).

Flood early warning systems are one of the most effective ways to minimize the loss of life and property. It is very important to have a reliable flood forecasting system as a basis for establishing a reliable early warning system, which can be transmitted down to the community in order to minimize the impact of flood disasters. Precipitation is highly variable in both space and time and is an important input in rainfall runoff modelling. The amount of rainfall and its spatial distribution are important factors in meteorology, climatology and hydrology. Accurate rainfall estimations are essential for timely flood forecasting and warning. In many regions, operational flood forecasting has traditionally relied upon a dense network of rain gauges or ground-based rainfall measuring radars that report in real time. Flood forecasting in basins with sparse or non-existent rain gauges poses an additional challenge (Shrestha et al, 2013). In such areas, satellite rainfall estimates (SRE) could provide information on rainfall occurrence, amount, and distribution (Adler et al. 2003; Hong et al. 2007; Shrestha et al. 2008 a,b) and can be used for hydrological modelling to predict floods.

2. STUDY AREA

Sudan is a country with a total area of 1,861,484 km^2 and is bordered by the Central African Republic, Chad, Egypt, Eritrea, Ethiopia, Libya and South Sudan. Covering two physiographic regions: the Arid and Semi-arid regions. Water-induced disasters are very prevalent and annually many lives are lost and property worth millions of dollars is destroyed.
Sudan has relatively few ground-based rain gauges, on average one gauge per 5170 km^2. Figure 1 shows the Sudan territorial area along with some rain gauge stations.

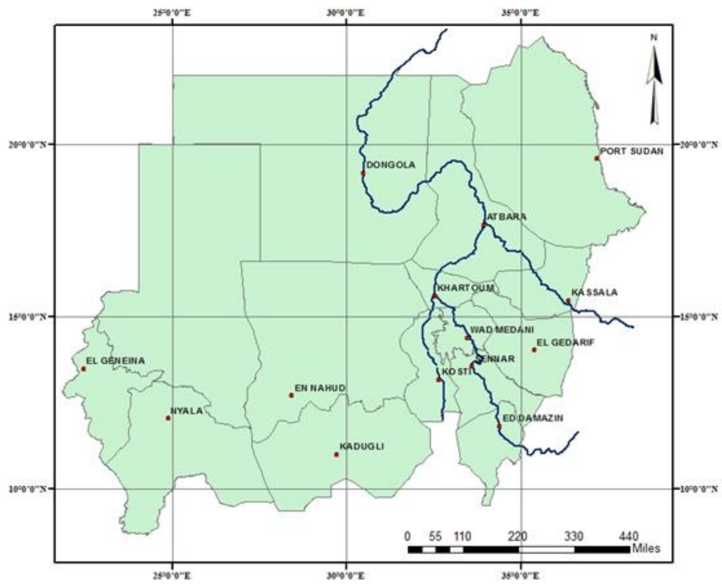

Figure 1: Sudan map and location of rain gauge stations

3. RAINFALL DATA

The monthly observed rainfall data for the period 2001 to 2010 from 14 gauge stations in Sudan were provided by the Sudan Meteorological Agency (SMA). The distribution of the rain gauges is shown in Figure (1). The density of rainfall gauge stations in southern part of the Sudan (south of Khartoum) is relatively high compared to northern part. However, the distribution is uneven and very sparse in the northern part where rainfall is very limited. Most stations are concentrated in urban areas where accessibility is easy.

Rainfall estimate algorithm (RFE) version 2.0 has been implemented by National Oceanic and Atmospheric Administration Climate Prediction Center (NOAA/ CPC). Created by Ping-Ping Xie, As January 1, 2001. The satellite RFE data used in this study are for the period 2001 to 2010 from (FEWS NET Data Portal)

4. METHODOLOGY

The satellite RFE data is downloaded in Band Interleaved by Line format (.bil) and converted to grid format in geographic projection. The grid is then projected to the Lambert Azimuthal projection using GeoStream Flow Model as extension under GIS environment. To obtain the rainfall value of a selected rain gauge, the raster value is extracted to point locations. These steps are done for each day from January 2001 to December 2010. The default unit for daily RFE is mm/day. For easy comparison, the rain rate in mm/day is converted to mm/month. This conversion is done by simply adding the daily rain rate for all the days in that particular month.

Statistical indices are used to compare and validate satellite RFE data with respect to gauge rainfall data. They help to understand the existing co-relationships, trends and error propagations between the two types of data sets. In this study, the statistical indices used are grouped into categorical verification statistics and continuous verification statistics.

4.1. Categorical Verification Statistics

Categorical verification statistics measure the correspondence between the estimated and observed occurrence of events and is a qualitative indicator. Most are based on a (2 x 2) contingency table of yes/no events, such as rain/no rain (Hojjat, S., 2010), as shown in Table (1) below. The probability of detection (POD) measures the fraction of observed events that was diagnosed correctly and is sometimes called the 'hit rate'. The false alarm ratio (FAR) gives the fraction of diagnosed events that were actually none vents (Ebert et al. 2007). The POD and FAR should always be used together.

Table 1: (2 x 2) contingency table

		Observed	
		Yes	**No**
Estimated	**Yes**	Hits	False Alams
	No	Misses	Correct Negatives

In rain/no rain contingency table shown above, the off-diagonal elements in the table characterize the error. The elements in the table (hits, misses, false alarms, correct negatives) give the joint distribution of events. In this table, hits represent correctly estimated rain events, where both satellite estimates and rain gauges show rain, false alarms represent when rain was estimated by satellite but did not occur on the ground and misses represent when rain was not estimated by satellite but did occur on the ground, and correct negatives represent correctly estimated no rain events. The contingency table is a useful way to see what types of errors are being made. A perfect estimate system would produce only hits and correct negatives and no misses or false alarms. Basic statistics are used to provide information on rain identification through contingency tables taken together with conditional rain rates (0 or 1 mm/day rain/no rain thresholds). This type of table was used to measure the skill of the rainfall estimations in pinpointing rain where rain was observed on the ground.

The Accuracy score shows the overall fraction of correct estimates. The range of this score is 0 to 1 and perfect score is 1. This score is simple and intuitive.

$$Accuracy = \frac{Hits + Correct\ Negative}{Total} \quad \ldots\ldots(1)$$

The probability of detection (POD) is sensitive to hits, but ignores false alarms. It is very sensitive to the climatology of the region and is good for rare events. It ranges from 0 to 1; the perfect score is 1.

$$POD = \frac{Hits}{Hits + Misses} \quad \ldots\ldots(2)$$

The false alarm ratio (FAR) is sensitive to false alarms, but ignores misses. It is very sensitive to the climatological frequency of the event and should be used in conjunction with the probability of detection. It ranges from 0 to 1; the perfect score is 0.

$$FAR = \frac{False\ Alams}{Hits + False\ Alams} \quad \ldots\ldots(3)$$

The Critical Success Index (CSI) measures the fraction of observed and/or estimated events that were correctly predicted, adjusted for hits associated with random chance. It is sensitive to hits. It penalizes both misses and false alarms in the same way and thus does not distinguish the source of estimated error. It ranges from -1/3 to 1; the perfect score is 1, 0 indicates no skill.

$$CSI = \frac{Hits - Hits_{random}}{Hits + Misses + Fals\ Alams - Hits_{random}} \quad \ldots\ldots(4)$$

$$Hits_{random} = \frac{(Hits + Misses)(Hits + Fals\ Alams)}{Total} \quad \ldots\ldots(5)$$

4.2. Continuous Verification Statistics:

Continuous verification statistics measure the accuracy of a continuous variable such as rain amount or intensity. These are the most commonly used statistics in validating satellite-based estimates; many people are familiar with them and find them easy to estimate.

The mean error (bias) measures the average difference between the estimated and observed values averaged over the data set. The range of mean error is minus infinity to infinity and the perfect score is zero.

$$ME(bias) = \frac{1}{N}\sum_{i=1}^{N}(S_i - G_i) \quad \ldots\ldots(6)$$

The root mean square error (RMSE) also measures the average error magnitude, but gives greater weight to larger errors. The range of RMSE is zero to infinity and the perfect score is zero. (Vila and Lima 2006).

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(S_i - G_i)^2} \quad \ldots\ldots(7)$$

The multiplicative bias (Mbias) is the ratio of estimated to observed rainfall values.

$$Mbias = \frac{\frac{1}{N}\sum_{i=1}^{N} S_i}{\frac{1}{N}\sum_{i=1}^{N} G_i} \quad \ldots\ldots(8)$$

The correlation coefficient (r) is measures the strength and the direction of a linear relationship between satellite RFE and observed gauge data.

$$r = \frac{\sum_{i=1}^{N}(S_i - \bar{S})(G_i - \bar{G})}{\sqrt{\sum_{i=1}^{N}(S_i - \bar{S})^2}\sqrt{\sum_{i=1}^{N}(G_i - \bar{G})^2}} \quad \ldots\ldots(9)$$

The percent of estimation (% of estimation) is the percentage of deference between RFE and observed to the observed data.

$$\%\ of\ Estimation = \frac{(\bar{S} - \bar{G})}{\bar{G}} * 100 \quad \ldots\ldots(10)$$

Where, S_i is the satellite RFE value at time i, G_i is the observed ground rain gauge value at time i, N is the number of observed samples, and $\bar{S}$ and $\bar{G}$ are the corresponding average values.

5. APPLICATION AND RESULTS

The two statistical groups of measures were applied to satellite derived monthly rainfall estimate (RFE) data over the period of 2001 to 2010 and the gauge rainfall data of 14 rain gauge stations during the rainy season across the Sudan. The results estimated for all stations are summarized in Table (2).

Table 2: Results of statistical measures

Statistical indices	Continuous Verification				Categorical Verification				% of estimation
Stations	r	bias	RMSE	M bias	Acc	POD	FAR	CSI	
Kadugli	0.60	-14.34	50.16	0.87	1.00	1.00	0.00	1.00	-12.6
Nyala	0.70	-13.57	48.37	0.81	0.93	1.00	0.07	0.93	-19.2
Eddamazin	0.82	-21.46	48.33	0.81	1.00	1.00	0.00	1.00	-18.8
ENahud	0.85	-10.37	35.49	0.83	0.95	1.00	0.05	0.95	-16.6
Elgeneina	0.86	-19.25	57.76	0.77	0.95	1.00	0.05	0.95	-22.9
Kosti	0.84	-7.47	37.71	0.88	0.98	1.00	0.02	0.98	-11.8
Sinnar	0.85	-11.09	36.95	0.83	1.00	1.00	0.00	1.00	-17.0
Wadmedani	0.65	-25.42	58.18	0.62	0.95	0.98	0.03	0.95	-37.6
Khartoum	0.78	-1.50	19.78	0.92	0.88	0.89	0.05	0.86	-7.5
Gedaref	0.87	-12.84	46.92	0.88	0.98	1.00	0.02	0.98	-12.4
Port Sudan	0.50	-0.71	8.11	0.75	0.75	0.83	0.57	0.40	-24.8
Kasala	0.80	4.19	26.48	1.11	0.90	0.96	0.07	0.90	11.3
Dongola	0.90	0.17	1.74	1.18	0.77	0.75	0.67	0.30	18.2
Atbara	0.60	1.46	12.84	1.24	0.88	0.86	0.11	0.78	23.6
Average	0.76	-9.44	34.92	0.89	0.92	0.95	0.12	0.86	-10.6

The results as shown in table (2) show variation from station to the other depending on their geographic location and variability of rainfall. The average correlation coefficient between observed and estimated rainfall is 76%. The range of mean error (bias) value is -25.42 to 4.19 with best value at Dongola station (0.17). Positive Mean Error (ME) indicates that the satellite RFE values are higher than the observed rainfall values; while negative Mean Error indicates areas where the satellite RFE values are lower than the observed rainfall values. Over the whole study area, negative ME are report to the Southern part of the Sudan and positive values reported to the Northern part. The range for root mean square error value is between 58.18 and 1.74 with an average value of 34.92. The estimated average probability of detection (POD) is 0.95, which is quite high indicating excellent detection rate by satellites while the estimated average false alarm ratio (FAR) of 0.12 relatively low.

The percentage of underestimations for each station are shown in table (2). Positive value indicates over estimation with average value (17.7%) and negative value indicates underestimation with average value (18.3%), with an overall average of (10.6%) under estimation.

The values of the total monthly observed and satellite RFE during rainy season for 2010 are given in Table (3) and Figure (2 a, b). Most of the rain falls during the rainy season (June to September) which constitutes 87% of the observed rainfall and 86% of the satellite RFE over the year. It can be seen that the satellite RFE values are lower than the observed values i.e. under estimation.

Table 3: Satellite RFE and Observed rainfall season 2010

Months	May		Jun		Jul		Aug		Sep		Oct	
Station	RFE	Obs	RFE	Obs	RFE	Obs	RFE	Obs	RFE	Obs	RFE	Obs
Port Sudan	0	0	0	13	6	5	5	6	0	0	2	0
Atbara	0	0	0	0	4	5	23	12	4	2	0	0
Dongola	0	0	0	0	8	11	6	4	0	0	2	0
Kasala	7	5	12	4	57	41	63	50	96	69	1	0
Khartoum	1	0	5	1	18	14	47	42	23	18	2	1
Wadmedani	15	18	22	30	44	85	71	89	22	32	12	7
Gedaref	13	15	65	76	133	218	177	208	98	95	9	35
Sinnar	10	21	35	72	83	148	80	144	29	54	35	21
Kosti	16	1.2	43	53	97	100	114	150	35	19	25	9.7
Elgeneina	0	0	19	48	183	230	129	179	80	100	24	38
ENahud	1	0	35	8	70	58	93	103	42	39	34	46
Eddamazin	36	73.5	123	212	120	91.5	154	105	115	183	56	55
Nyala	0	0	33	62	58	97	56	44	85	135	50	40
Kadugli	74	50	93	96	96	142	129	265	122	78	81	195
Average	12.4	13.1	34.6	48.2	69.8	89	81.9	100	53.6	58.9	23.8	32

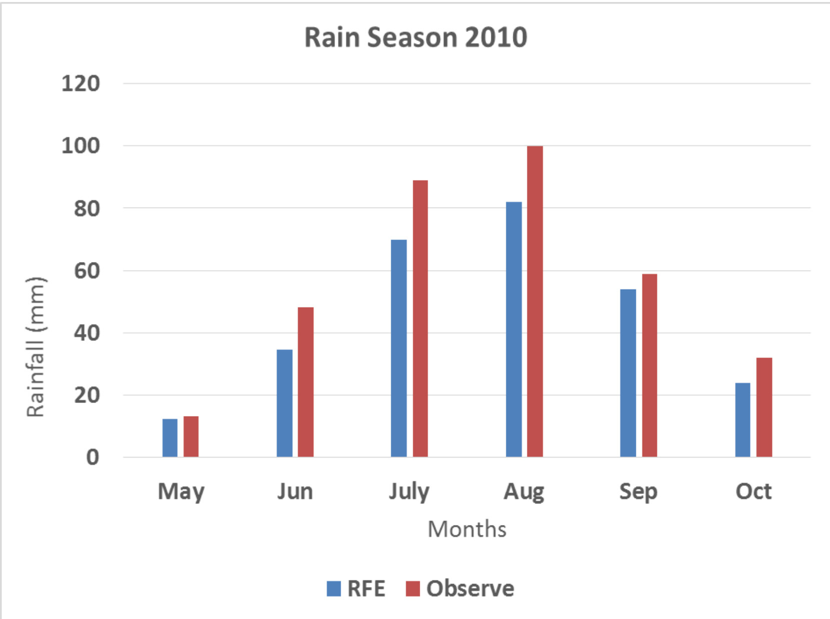

Figure 2a: Satellite RFE and Observed rainfall season 2010

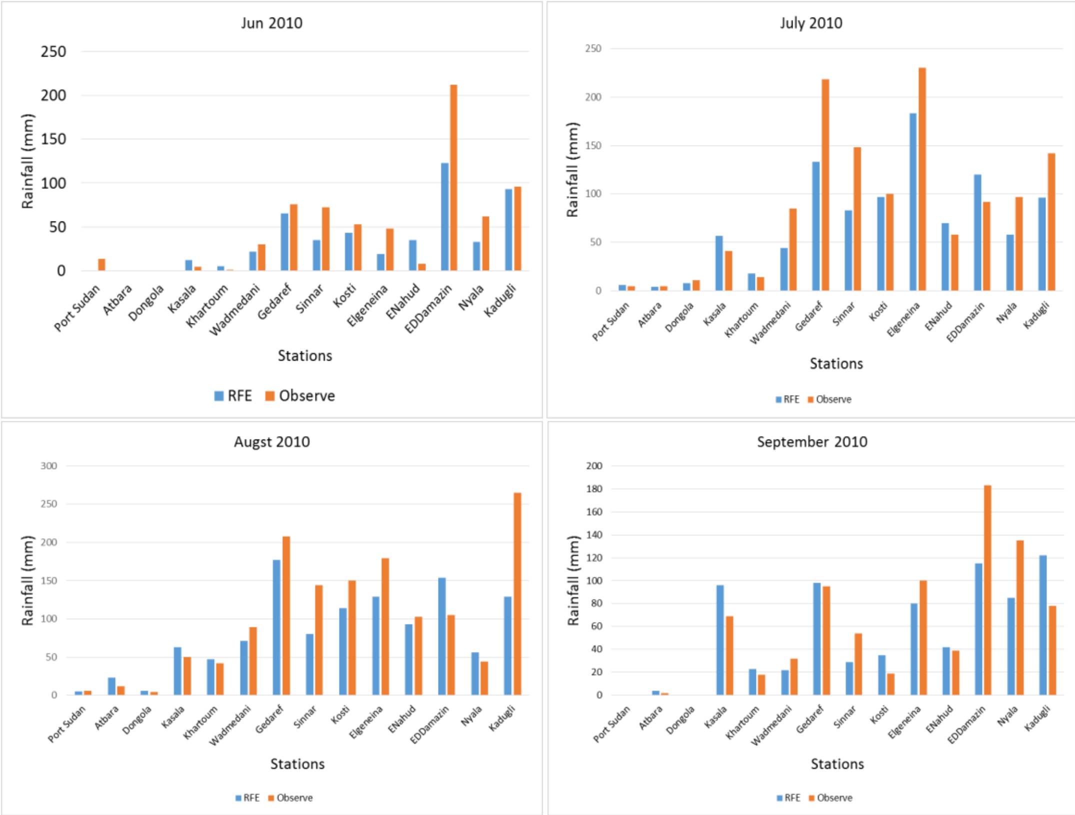

Figure 2b: RFE and Observed rainfall for the stations (Jun-Sep) 2010

6. SUMMARY AND CONCLUSIONS

The general objective of this study evaluate the reliability of the Satellite CPC_RFE 2.0 data by using monthly rain gauges data was achieved with categorical and continuous statistical analysis of 14 stations covering most area of the Sudan during the rainy season.

Analysis on the time series of monthly rainfall of satellite RFE and observed data in different location in the Study area revealed that all satellite RFE data are significantly correlated with gauge rainfall data giving mean correlation coefficients of 76%. The different comparisons show that the CPC_RFE2.0 underestimates the monthly rainfall for the observed rain gauges over the whole country except Dongola, Kasala and Atbara. The correlation coefficient in the southern part of Sudan is better than northern part. In addition, continuous validations show that the RFE performed well in different climate zones, with considerably low mean error (ME) and root mean square error (RMSE) scores. For the dichotomous validation the probability of detection (POD) values is above 0.86 while the false alarm rate (FAR) is low than 0.07 except at Dongola, Port Sudan and Atbara stations. Satellite NOAA CPC RFE 2.0 performed well in the estimation and monitoring of rainfall over the Sudan and can be used to analyze the precipitation pattern, delineate the flash flood hazard areas and produce estimated discharges for rain water harvesting and irrigation projects.

7. REFERENCES

1. Adler, RF., Chang, GJ., Ferraro, R., Xie, P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P. and Nelkin, E. (2003), *The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979-present),* J. Hydrometeorol. 4: 1147–1167.
2. Bajracharya, M.S. Shrestha and A.B. Shrestha. (2014), *Assessment of high resolution satellite rainfall estimation products in a streamflow model for flood prediction in the Bagmati basin, Nepal*, International Centre for Integrated Mountain Development (ICIMOD), Kathmandu, Nepal.
3. Ebert, EE., Janowiak, JE., Kidd, C. (2007), *Comparison of near-real-time precipitation estimates from satellite observations and numerical models*, Bull. Amer. Meteorol. Soc, 88(1): 47–64.
4. Hojjat, S., (2010), *Comparing satellite derived rainfall with ground based radar for North-western Europe*, Master thesis, International Institute for Geo-information Science and Earth Observation.
5. Hong, Y., Adler, RF., Negri, A., Huffman, GJ. (2007), *Flood and landslide applications of near real-time satellite rainfall estimation*, Nat. Hazards, 43: 285–294.
6. Levizzani, V., Bauer, P., Turk, F. Joseph (2007), *Measuring Precipitation from Space: EURAINSAT and the Future, published by Springer*, the Netherlands, ISBN 978-1-4020-5834-9, pp 345-356.
7. Shrestha, MS., Artan, GA., Bajracharya, SR., Sharma, RR. (2008a), *Applying satellite-based rainfall estimates for streamflow modelling in the Bagmati Basin*, Nepal, J. Flood Risk Management, 1: 89–99.
8. Shrestha, MS., Bajracharya, SR., Mool, P. (2008b), *Satellite rainfall estimation in the Hindu Kush-Himalayan Region*, Kathmandu, Nepal: ICIMOD.
9. Shrestha, MS., Rajbhandari, R., Bajracharya, SR. (2013), *Validation of NOAA CPC_RFE satellite-based rainfall estimates in the central Himalayas*, ICIMOD Working Paper 2013/5. Kathmandu: ICIMOD.
10. Vila, D; Lima, A. (2006), *Satellite rainfall estimation over South America: The hydro estimator technique*, In WW Grabowski (ed) 14th International Conference on Clouds and Precipitation, Bologna, Italy, 18–23 July 2004. Amsterdam, Netherlands: Elsevier.

AUTHORS BIOGRAPHY

Author[1]: Aboubaker Ahmed Osman Ali, MSc in Civil engineering (Water Resources), Khartoum University, faculty of engineering, working lecture in Nile Valley University faculty of engineering

Author[2] : Babiker Ibrahim Barsi, Associate professor in civil Engineering, University of Khartoum

Author[3] : Kamaleldin Elsiddig Bashar, Professor of water Resources, faculty of Engineering, Omdurman Islamic University