Estimating Areal Rainfall over the Lake Victoria and its Basin using Ground-based and Satellite Data

Michael Kizza[1], Ida Westerberg[2], Henry K. Ntale[3], Allan Rodhe[4]

[1]Assistant Lecturer, Faculty of Technology, Makerere University, P. O. Box 7062, Kampala, Uganda
Corresponding author email: mkizza@tech.mak.ac.ug
[2]Graduate student, Department of Earth Sciences, Uppsala University, Villavägen 16, SE-752 36 Uppsala, Sweden
[3]Senior Researcher, Faculty of Technology, Makerere University, P. O. Box 7062, Kampala, Uganda
[4]Professor, Department of Earth Sciences, Uppsala University, Villavägen 16, SE-752 36 Uppsala, Sweden

ABSTRACT

A gridded monthly rainfall dataset having a spatial resolution of 2 km was derived for Lake Victoria basin. Such a dataset is useful for hydrological modelling in the basin aimed at resource utilisation and also for estimation of catchment flow to Lake Victoria. Using cross-validation we showed that universal kriging interpolation was slightly better than inverse distance weighting for the basin. The rainfall patterns in the dataset were assessed and shown to reflect the expected patterns. Rainfall over the lake was estimated from mean observed rainfall over the land area using a correlation between lake and land area derived for two satellite products. The TRMM 3B43 product showed an enhancement of lake rainfall over basin rainfall of 25% while the PERSIANN product gave a much higher enhancement of up to 80%. Regression based on calendar months resulted in a slight increase in the estimated mean areal rainfall over the lake of 1.0% for TRMM 3B43 and 2.2% for PERSIANN compared to using a single regression equation for all months. The usefulness and limitations of the two satellite products in estimation of lake rainfall are discussed.

Keywords: Arial rainfall, Lake Victoria Basin, Satellite data

1.0 INTRODUCTION

With a surface area of 68,800 km^2, Lake Victoria is the largest freshwater lake in the tropics and the second largest lake in the world. The lake and its basin support over 30 million people, who directly or indirectly depend for their livelihoods, drawing on services like fishing, transport, agriculture, water supply, tourism and hydropower. Being one of the sources of the Nile, the lake is a major supplier for the water needs of Sudan and Egypt. Additionally, the lake controls the hydrology of the Sudd wetlands in southern Sudan whose areal coverage averages about 35,000 km^2 but can be as much as 130,000 km^2 mainly due to high flow in the Nile. However, 80% of the input into the lake's water balance is rainfall over its surface, leading some researchers to describe it as 'atmosphere controlled' (Yin and Nicholson, 1998). This essentially means that the variability of rainfall over the lake plays a key role in the fluctuation of the lake levels. Historically, the lake levels have exhibited large and rapid changes in response to rainfall anomalies over the last century (Fraedrich, 1972). Changes in the lake basin rainfall regime have far reaching ecological, environmental, hydrological and socio-economic effects. Fluctuations in the level of Lake Victoria have generated controversy surrounding the potential returns of new hydro-power installations. Within the lake basin, a combination of factors is exerting serious stress on the available water resources. Rapid population growth rates, estimated at over 3%, high poverty rates as well as developmental activities are causing serious deforestation. Reclamation of wetlands, siltation of rivers and other water bodies due to erosion from the now bare soils and increase in point and non-point pollution, are serious problems. There are two major problems related to the assessment of spatial variability of rainfall in the Lake Victoria basin. One problem is related to the differences in distribution of rain gauges around the lake (Figure 1). The quality of the rainfall data also varies considerably depending on data source and period analysed. The second problem is the absence of reliable long term rain gauges on the lake surface.

This study had two objectives. The first objective was to use the available rain gauge data in the basin to derive a spatially detailed gridded monthly rainfall dataset for the lake basin using geo-statistical techniques. Such a detailed dataset is useful as an input to hydrological studies assessment of discharges of the rivers that flow into Lake Victoria. The second objective was to improve estimates of lake surface rainfall using remotely sensed datasets. Two satellite products, namely, PERSIANN and TRMM 3B43, which have been shown to work considerably well for the region (Dinku et al., 2007; Asadullah et al., 2008), were used to derive regression equations between lake and land rainfall in the basin. Finally, we produced time series of areal mean rainfall for the main tributaries into Lake Victoria.

2.0 STUDY AREA AND DATA

2.1 The Lake Victoria basin

Lake Victoria is located between latitudes 0°20'N-3°S and longitudes 31°40'E-34°53'E (Figure 1). The basin has an area of 194,000km^2 and the lake surface area is about 68,800 km^2 or 35% of the basin. The lake surface is shared between Kenya (6%), Uganda (43%) and Tanzania (51%) while its basin includes parts of Burundi and Rwanda. The altitude of the lake surface is about 1,135m above mean sea level (a.s.l). The basin is made of a series of stepped plateaus with an average elevation of 2,700m a.s.l. but rising to 4,000m a.s.l. or more in the highland areas. The lake is relatively shallow with an average depth of 45 m and maximum depth of 92 m. The main river flowing into the lake is the Kagera which, together with four other tributaries (Nzoia, Yala, and Sondu and Awach-Kaboun), contribute about 50% of the flow.

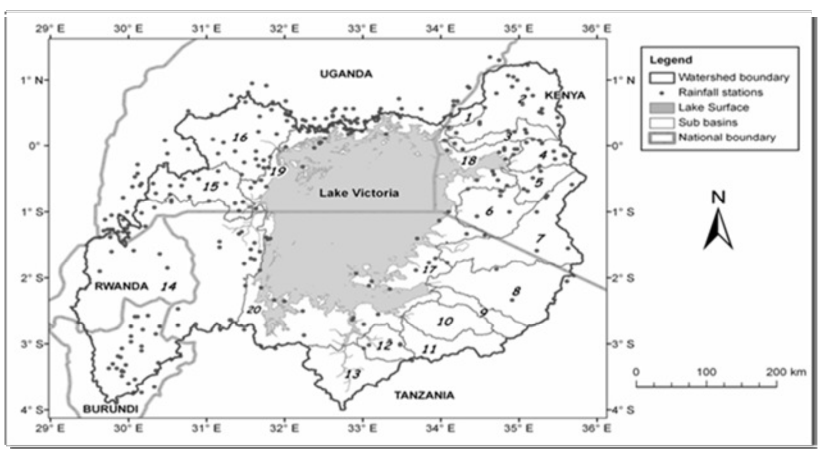

Figure 1: Lake Victoria basin; the numbers 1-20 are the sub-basins.

2.2 Rainfall regime

The general climate of Lake Victoria basin ranges from a modified equatorial type with substantial rainfall occurring throughout the year, particularly over the lake and its vicinity, to a semiarid type characterized by intermittent droughts over some areas located even within short distances from the lakeshore. The diurnal, seasonal and inter-annual variability of the basin climate (and East Africa generally) results from a complex interaction between the Inter-tropical Convergence Zone (ITCZ), El Nino/Southern Oscillation (ENSO), Quasi-biennial Oscillation (QBO), large scale monsoonal winds, meso-scale circulations and extra-tropical weather systems (Anyah *et al*., 2006). The wind and pressure patterns that govern the region's climate include three principal air streams namely; the Congo airstream with westerly and south-westerly winds, the southeast monsoon and the northeast monsoon. The north-south migration of the ITCZ results in a predominantly bimodal seasonal rainfall distribution for the basin; the March–May rainfall period (long rains) and the October–December rainfall period (short rains). Inter-annual variability corresponds to the ENSO variability. El Niño years are usually associated with above normal rainfall amounts in the short rainfall season in most of the region. Meso-scale circulations due to orography, lake surface temperature and other factors have also been shown to significantly amplify the rainfall over the lake compared to surrounding land (Anyah et al., 2006).

2.3 Data

The rain gauge data used in this study were collected from various sources including the hydro-meteorological database of the meteorological departments in Kenya, Tanzania and Uganda as well as from our correspondence with other researchers in the region. The data format was either as the raw daily values or aggregated monthly values. From these sources, a dataset of over 1000 daily rainfall stations was created. Out of these, 362 stations having at least 2 years of record were selected for further analysis.

Two satellite rainfall products, namely TRMM 3B43 and PERSIANN were used for estimating lake rainfall on the basis of their good prior performance for the region (Dinku et al., 2007; Asadullah et al., 2008). TRMM 3B43 is a monthly satellite rainfall data product of the NASA Goddard Space Flight Centre, USA. The TRMM 3B43 data used for the current study covers the period from January 1998 to December 2009. The PERSIANN product, from University of California, Irvine, USA, uses an artificial neural network algorithm to combine the infrared and passive microwave data. The PERSIANN data used for this study cover the period from March 2000 to December 2009. Both TRMM 3B43 and PERSIANN have a spatial resolution of 0.25^{o} x 0.25^{o}.

3.0 METHODS

The study was divided into 3 parts; (i) data quality control and filling of missing data, (ii) spatial interpolation of gauge data, and (iii) regression analysis with satellite data to estimate lake rainfall. The data used in the analysis covered the period 1960-2005.

3.1 Data quality control

Quality control was performed at both daily and monthly time scales and potentially erroneous data were flagged using automatic quality control routines and visual inspection. Flagged data were further investigated and removed if deemed necessary. Automatic routines similar to the ones developed by Westerberg et al. (2010) for a Honduran basin were used to check for; (1) too frequently occurring data, (2) sequences of too low or high data, (3) unusually dry or wet months, and (4) outliers. During quality control, three types of decisions had to be made depending on the evidence gathered namely; (1) retain the data and station, (2) remove the erroneous data from the station data series or (3) remove the entire station from the database. The decision to remove data was subjective but the previous analysis provided a sound basis for this decision.

3.2 Estimation of missing values

Missing values in the rainfall dataset were filled using the coefficient-of-correlation-weighting (CCW) method which was shown to work well by Teegavarapu and Chandramouli (2005). CCWM uses the correlation coefficient instead of inverse squared distance to compute the weighting factor:

$$\theta_m = \frac{\sum_{i=1}^{n} \theta_i R_{mi}}{\sum_{i=1}^{n} R_{mi}} \tag{1}$$

Where θ_m is the missing rainfall value to be interpolated, R_{mi} is the correlation coefficient between the data at station *m* with the missing value and the i^{th} surrounding station, θ_i is the rainfall value at station *i* and *n* is the number of surrounding stations.

3.3 Spatial interpolation

Many different spatial interpolation methods exist. Here two methods of different complexity that are both based on weighted linear combinations of the point rainfall data were selected; IDW and universal kriging (UK) with coordinate base functions. Despite its simplicity, IDW can produce results that are comparable to the more computationally demanding kriging if the sampling density is high. The advantage of using kriging is in situations where the sampling density is low, especially if the spatial dependence between the variables is high. IDW is a straight-forward, deterministic interpolation method that assigns weights based on the inverse of the squared distance to every data point located within a given search radius, centred on the point to be estimated. UK is a member of techniques grouped as geo-statistical estimation methods in which the data to be interpolated are treated as the outcome of a random function model (Isaaks and Srivastava, 1989). In the random function model, the pattern of spatial continuity is described through the sample semi-variogram which is estimated from the data. UK was performed as block kriging (Isaaks and Srivastava, 1989) with a grid size of 2,000 m and a discretisation of 100 points per block.

3.4 Estimation of lake rainfall

A three step approach to estimate lake rainfall was used in this study based on the two satellite rainfall products, TRMM 3B43 and PERSIANN. First, an assessment was carried out on how well the two products reproduced areal rainfall over the land part of the basin as estimated using interpolation of rain gauge data. The second step was to derive the relationship between areal lake rainfall and areal rainfall in the surrounding catchment by regression analysis for each calendar month with satellite estimated lake rainfall as dependent variable and satellite estimated rainfall over the land as independent variable. The resulting equations were then applied to derive the areal lake rainfall series for the period 1960 to 2004 from the areal rainfall estimates over the land part of the basin from interpolation of rain gauge data.

4.0 RESULTS AND DISCUSSION

4.1 Data quality control

A total of 47 stations (13%) out of 362 stations were dropped during the quality control process. The 315 remaining stations represent a network density of 620km^2/gauge over the lake basin and 830km^2/gauge if the lake area is included. Furthermore, a total of 42 (12%) stations had their data edited by removing some data values that were deemed unreliable. Of these, 31 stations were edited because of outliers at daily time step, 6 stations because of repeated values and 5 stations because of having sequences of abnormally high or low values as reflected by outliers in the monthly and annual timescale but not at daily timescale. The number of data removed from each affected station in the dataset ranged between 1 day and 2,920 days (8 years).

4.2 Estimation of missing data

Gap filling also resulted in an increase in the mean data length from 14 to 19 years for the period 1960-1984 and from 7 to 10 years for the period 1985-2004. It is important to note that the gap filling CCWM process was robust and preserves most of the key statistics of the dataset. The mean annual rainfall for the entire dataset (1960-2004) before and after gap filling changed from 1268mm to 1262mm respectively while the coefficient of variation changed from 0.188 to 0.193 respectively.

4.3 Spatial interpolation

4.3.1 Comparison of the interpolation methods

Inverse distance weighting (ID) and universal kriging (UK) methods gave similar results during cross validation (Table 1). Most of the statistical measures were of the same order though it might be said that, overall, UK performed slightly better. As expected the statistics for 1960-1984 were better than the statistics for 1985-2004 because, on average, there were more stations for 1960-1984 and therefore more data points to interpolate.

Table 1: Statistics of results based on interpolation of monthly rainfall totals using the two methods and for different time periods

Interpolation method	IDW			UK		
Time period	1960-1984	1985-2004	1960-2004	1960-1984	1985-2004	1960-2004
Coefficient of correlation	0.61	0.4	0.52	0.61	0.41	0.52
Mean error (mm/month)	-1.5	0.44	-0.64	-0.17	-0.1	-0.14
Mean absolute error (mm/month)	32.5	33.3	32.8	32.4	32.4	32.4
Root mean square error (mm/month)	46.2	48.3	47.1	45.7	47	46.3

4.3.2 Spatial rainfall characteristics

For illustrative purposes, mean monthly rainfall for the entire analysis period (1960-2004) estimated by the UK method were plotted in Figure 3, to show the seasonal variation of rainfall in the basin. The rainfall field over the lake was far from accurate, and is the subject of discussion in section 4.4. However, the rainfall patterns produced by the two interpolation methods are consistent with expected variations in the land areas. The mean annual rainfall patterns resulting from interpolation with UK and ID were also similar (Figure 2). From the results of UK interpolation, the land part of the basin received an annual average of 1227mm with a maximum of 180mm in April and a minimum of 45mm in July.

4.4 Estimation of lake rainfall

4.4.1 Comparison of TRMM 3B43 and PERSIANN estimates with ground data

The bias for TRMM 3B43 (-7.0% for UK and -6.4% for ID) was higher than the values for PERSIANN (-1.8% for UK and -1.2% for ID). The bias in TRMM 3B43 estimates is consistent with those quoted for Australia of about 6% (Huffman et al., 2007) while the bias in PERSIANN estimates is consistent with values quoted for the region by Asadullah et al. (2008). The correlation coefficients for TRMM 3B43 (0.90 for UK and 0.88 for ID) were higher than those for PERSIANN (0.75 for UK and 0.71 for ID). The basic statistics of mean, maximum, minimum and standard deviation were also quite similar (Table 2).

Table 2: Statistics of monthly rainfall for the period 2001-2004 over the land part of the Lake Victoria basin for the satellite products and using the two interpolation methods

Statistic	TRMM 3B43	PERSIANN	UK	ID
Mean (mm/month)	94	100	101	101
Maximum (mm/month)	194	211	209	210
Minimum (mm/month)	22	11	23	27
Standard deviation (mm/month)	46	53	51	48

4.4.2 Regression of lake rainfall

Linear regression equations for the satellite products were derived to relate the rainfall over the Lake Victoria surface (dependent variable) to rainfall over the land part of the basin (independent variable). The results for TRMM 3B43 and PERSIANN are shown in (Table 3).

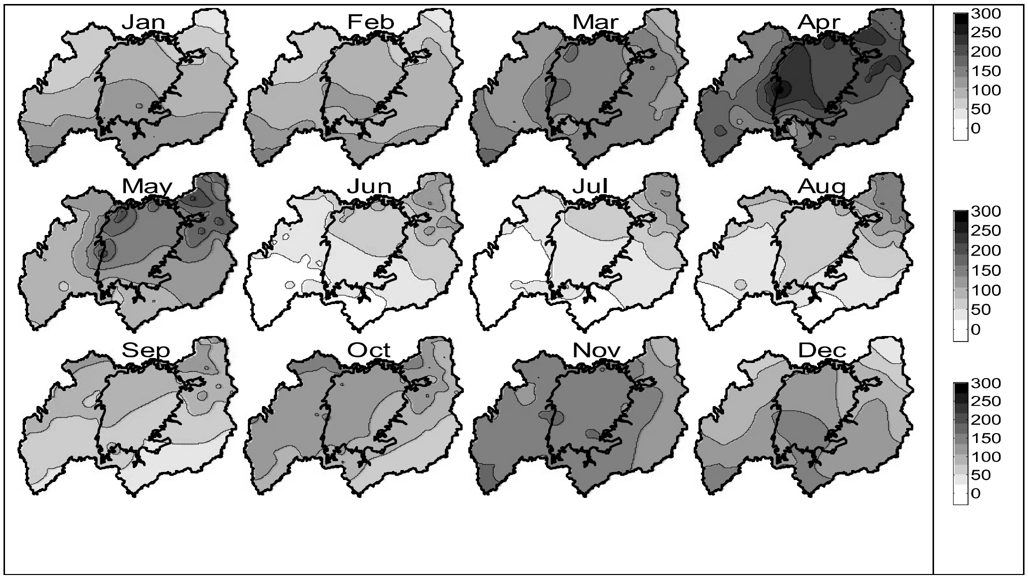

Figure 2: Mean monthly rainfall (mm/year) for UK for 1960-2004

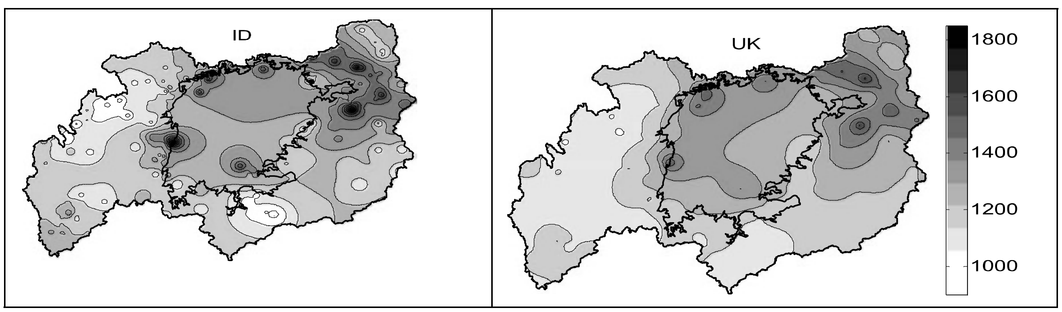

Figure 3: Spatial variations of the mean annual rainfall (mm/year) interpolated by ID (left) and UK (right) methods.

Table 3: Regression coefficients and performance measures for monthly lake rainfall against monthly land rainfall for the TRMM 3B43 (1998-2009) and PERSIANN (2001-2009) satellite products. The equations are of the form $y = a + bx$. The column entitled 'ratio' gives the ratio between lake rainfall and land rainfall. Variable *a* and RMSE have units of mm/month)

TRMM 3B43					PERSIANN			
a	*b*	R^2	RMSE		*a*	*b*	R^2	RMSE
-4.78	1.32	0.903	24.5		16.74	1.64	0.815	42.4

The TRMM 3B43 satellite product was used to estimate lake rainfall with help of the regression equation established before. On average, the enhancement of lake rainfall over land rainfall was 25% (Table 4). The TRMM 3B43 results were more consistent with other results quoted in literature (Fraedrich, 1972; Hastenrath and Kutzbach, 1983; Ba and Nicholson, 1998; Yin and Nicholson, 1998). Hastenrath and Kutzbach (1983) computed an enhancement of 25% of lake rainfall compared to the rainfall in the surrounding catchment. Fraedrich (1992) used a simple climatological model to estimate an annual rainfall of 1,585mm over the lake compared to about 1,000 mm over the land part of the catchment, or an enhancement of 59%. Ba and Nicholson (1998) computed a relationship between satellite-derived cold cloud indices and rainfall measurements to show an enhancement of 27% of lake rainfall over land rainfall.

Table 4: Mean monthly lake rainfall estimates from UK derived areal rainfall over land part of Lake Victoria. Case 1 is based on a single regression equation for all data while Case 2 is based on regression equations for each calendar month.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual (mm/year)
Land rain (UK)	93	91	136	185	127	52	48	64	76	105	138	111	1,227
Lake rain (TRMM 3B43)	115	113	171	232	159	63	58	81	96	133	174	136	1,531
Lake rain (PERSIANN)	165	163	235	311	220	101	94	123	141	188	238	191	2,170

5.0 CONCLUSIONS

The following conclusions are drawn from the above analysis

- With careful data quality assessment and gap filling, the ground-based rainfall records provide a good basis for analysing spatial rainfall variability over the land part of Lake Victoria basin
- While the two interpolation methods used are different conceptually, they provide comparable results for the Lake Victoria basin
- Both satellite products capture the enhancement of lake rainfall compared to land rainfall. The magnitude of the enhancement is different for the two satellite products. Application of this kind of analysis to other satellite rainfall products could throw more light on whether the magnitude of enhancement can converge to a uniform value.

6.0 REFERENCES

Anyah, R. O., Semazzi, F. H. M.and Xie, L. (2006), *Simulated physical mechanisms associated with climate variability over Lake Victoria in East Africa*, *Mon.* Wea. Rev., 134**,** 3588-3609.

Asadullah, A., Mcintyre, N.and Kigobe, M. (2008), *Evaluation of five satellite products for estimation of rainfall over Uganda*, Hydrological Sciences Journal/Journal des Sciences Hydrologiques*,* 53**,** 1137-1150.

Ba, M. B.and Nicholson, S. E. (1998), *Analysis of convective activity and its relationship to the rainfall over the Rift Valley Lakes of East Africa during 1983–90 using the meteosat infrared channel*. J. Appl. Meteorol., 37**,** 1250-1264.

Dinku, T., Ceccato, P., Grover-Kopec, E., Lemma, M., Connor, S.and Repelewski, C. (2007), *Validation of Satellite Rainfall Products over East Africa's Complex Topography*, International Journal of Remote Sensing*,* 28**,** 1503-1526.

Fraedrich, K. (1972), *A simple climatological model of the dynamics and energetics of the nocturnal circulation at Lake Victoria*, Quarterly Journal of the Royal Meteorological Society*,* 98**,** 322-335.

Hastenrath, S.and Kutzbach, J. E. (1983), *Paleoclimatic estimates from water and energy budgets of East African lakes*. Quaternary Research*,* 19**,** 141-153.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P.and Stocker, E. F. (2007) *The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales*. Journal of Hydrometeorology*,* 8**,** 38-55.

Isaaks, E. H.and Srivastava, R. M. (1989), *An introduction to applied geo-statistics*, Oxford University Press, New York.

Yin, X.and Nicholson, S. E. (1998), *The water balance of Lake Victoria*, Hydrological Sciences Journal*,* 43**,** 789-811.