Hydrology: Current Research

Research Article — Open Access

Estimation of Missing River Flow Data for Hydrologic Analysis: The Case of Great Ruaha River Catchment

Lusajo H Mfwango*, Catherine J Salim and Shija Kazumba

Department of Water Resources and Irrigation Engineering, Water Institute, Dar es Salaam, Tanzania

*Corresponding author: Lusajo H Mfwango, Department of Water Resources and Irrigation Engineering, Water Institute, P.O.Box 35059, Dar es Salaam, Tanzania, Tel: +63280009456; E-mail: lusajojo2002@yahoo.com

Abstract

Availability of data on hydrologic variables such as river flow is necessary for planning and management of water resources. Many developing countries Many River basins in developing countries has no complete dataset on river flow due to degradation of gauging stations gauging stations coupled with unsatisfactory data compilation unsatisfactory data compilation and storage procedures. Different methods are available to fill missing data; however, these methods differ in performance depending on the characteristics of initial data points. The purpose of this study was to fill the missing data in the Great Ruaha River by selection of best method. In this study, simple and multiple regression analysis, and recession methods have been employed to fill the gaps of missing river flow data on ten gauging stations of Great Ruaha River catchment. Performances of these methods were assessed using Nash-Sutcliffe efficiency, Root Mean Square Error and Mean Absolute Error. The results showed that, Multiple regressions are suitable over Linear regression method for missing data during the period of high flow, however selection of either method depends on the availability of data availability on independent variable. Recession method was found to be suitable for filling missing data during the period of low flow. Though these methods were useful in filling data, the challenge was that more than one method was required to estimate all the missing data at a gauging station. This is because, missing data at a given gauging station were experienced during dry and rain seasons.

Introduction

Great Ruaha River is the among the major driving force to the economy of Tanzania through various sectors such as hydropower production, domestic water supply, irrigated agriculture and fishing industry. It originates from Kipengere mountains in Southern central of Flowing through the usangu wetland and Ruaha National Park East into the Rufiji River which drains its water into Indian Ocean [1]. Agriculture is the most economic activity in the basin, it exploit to majority in the basin exploit large amount of water from the river results into many conflicts with other sector especially during the dry period. In additional to the aforementioned problem, climatic change also accelerates the drying of many tributaries of Great Ruaha river except for perennial rivers such as Mbarali, Kimani, Chimala which retain very minimal flow in the dry season [2].

Sustainability of water resources requires proper planning for development and management. However, without a complete data set on river flow the above-mentioned purposes cannot be achieved. Many gauging stations along Great Ruaha River has poor data quality characterized by short and long duration of missing data. Elshorbagy has outlined some of the factors which might attributed to the existence of data gaps, such factors include equipment failure, effects of natural phenomena such as flooding or mishandling of observed records by field personnel, or accidental loss of data files [3]. Various methods or techniques applicable for the infilling of data are given in the literature, these include, regression methods and Recession [4,5].

Despite the availability of numerous methods in filling of missing hydrological data, it is generally believed that no single method can be considered universally best . Each method has its own advantages and disadvantage depending on the characteristics of the data set. A method that fit some data point can be unsuited for a different set of data points, or if measured in different locations of the same surface.

However other factors as reported by DeSilva et al. such as s distances among gauges, aerial coverage of each gauge, length of gap, the season, the climatic region or the availability and data characteristics of the records has significant influences on hydrological data estimates [6].

The aim of this study is to estimate missing river flow data for hydrologic analysis using regression analysis and recession methods as well as comparing estimation performance of the selected methods.

Description of Study Area and Data Availability

The study was conducted at Great Ruaha catchment which is located in southwest of Tanzania between latitudes 5°30'' and 9°25'', South, and longitudes 33°55'' and 37°80'' East (Figure 1). The Great Ruaha River is among the most importants with a great impact to the economy of Tanzania. It is a main tributary (catchment area ≈ 68,000 sq km) of the Rufiji River which forms the largest drainage basin among the nine basins, covering an area of 177,000 sq. km of the Tanzania mainland [7].

Within the catchment other small rivers such as Mbarali, Kimani and Chimala from the highlands join Great Ruaha River flowing through Usangu plain and Ruaha National Park while serving as are major sources of water to the irrigated agriculture and wildlife

respectively. Thereafter together with little Ruaha River supplies water to Mtera hydropower plant and form Rufiji River downstream the power plant [2].

The area characterized with mean annuals the mean annual air temperature which varies from about 18°C, at higher attitudes to about 28°C in the lower and drier part of the basin. Rainfall pattern is unimodal (November to April), highly localized and spatially varied and strongly correlated with altitude. Usangu plain at the lower altitude, has rainfall ranging from 500-700 mm per annum while at the higher altitude up to about 1600 mm of rain [8].

For the purposes of this study the daily river flow data for eleven gauging stations within the catchment were obtained from Rufiji Basin Water Board (RBWB). Detailed information on the data is given in Table 1.

Gauging station	Main river	Period of record (month/year)	No. of years	Data availability (%)
Chitekero (IKA7A)	Chimala	01/1990-11/2014	24	88
salimwani (IKA8A)	Great Ruaha	01/1998-06/2014	16	92
Msembe (1KA59)	Great Ruaha	01/1994-11/2014	20	90
GNR (1KA9A)	Kimani	03/1991-11/2014	23	97
Igawa (1KA11A)	Mbarali	01/1998-09/2014	16	98
Ilongo (1KA15A)	Ndembera	01/1990-09/2014	24	91
Ihimbu (1KA21A)	Little Ruaha	06/1991-12/2014	23	90
Makalala (1KA32A)	Little Ruaha	01/1990-10/2014	24	92
Ndiuka (1KA2A)	Little Ruaha	01/1990-11/2014	24	67
mtandika (1KA37A)	Lukosi	01/1990-12/2014	24	86
Mtitu (1KA22)	Mtitu	01/1990-12/2014	24	100

Table 1: General Information of the gauging stations in Great Ruaha Catchment.
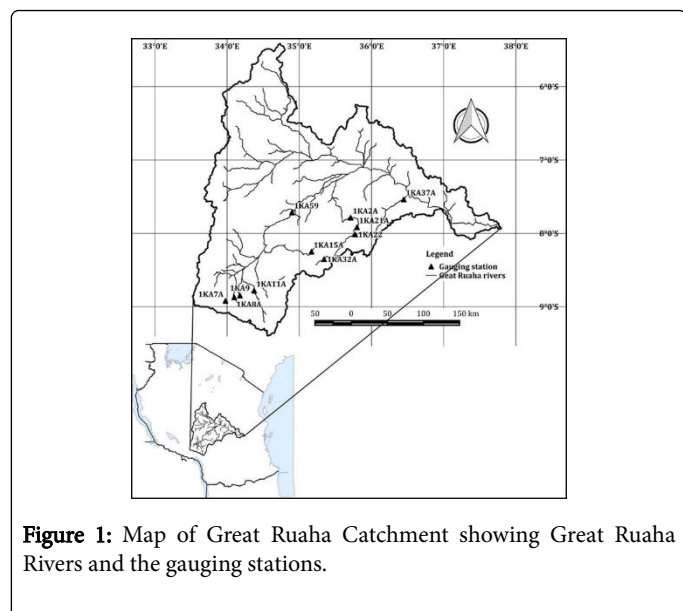
Figure 1: Map of Great Ruaha Catchment showing Great Ruaha Rivers and the gauging stations.

In some of these gauging stations, data for one day, months or even a year were missing. Among these gauging stations, Mtitu (1KA22) is the only gauging station with a complete dataset, while Ndiuka (1KA2A) which is located along Little Ruaha River has lowest data availability of 67%.

Methodology

In a process of filling missing data, all eleven gauging stations were involved. For regression methods, the selection of independent and depend variables were based on the following factors; the correlation coefficient between gauging stations, Data availability for the donor stations (independent variables) and location of the gauging stations within the catchment. Correlation Coefficient is an indicator for the strength of the relationship between two variables. Higher positive coefficients between variables indicate that estimates will be high or low when actual is high or low respectively giving evidence about the suitability of the method [6]. Hence the correlation coefficients were determined for 100 combinations of gauging stations. The gauge stations with strong correlation in consideration with other criteria were chosen to predict the missing data of the either stations.

The next step involved calibration process to develop an equation that describes the dataset. Elshorbagy et al. noted that in hydrological data (e.g., stream flows), annual or seasonal data might be independent, while monthly or weekly data of the same river have significant levels of autocorrelations [9]. Hence calibration was done by selecting gauging stations with a complete (without gap) dataset of five years for two (Linear regression analysis) or three (Multiple regression analysis) stations with strong correlation between them. The developed equation was used to fill missing data of the dependent variable during the period of high flow.

For Recession method, a base flow recession factor, which characterizes the behavior of low flow was derived from a simple exponential equation. This was done by averaging five recession factors

($α1$, $α2$, $α3$, $α4$ and $α5$) from five-year consecutive low flow dataset recorded at the gauging station. The value obtained was substituted to the recession equation to fill the missing data.

In order to test the accuracy of methods used in estimation of missing data, a gauging station (X) and gauging station/stations for which data are available, were selected and assumed that observations from X station are missing. Then using each method, observations for X station are estimated and compared with the actual observations.

The estimates obtained from each method were compared with observed records. The suitability of method is decided by how close the estimates and observed values are in a given time series. For this study, Nash Sutcliff efficiency, Root Mean Square error and Mean Absolute Error were used as criteria to estimate the closeness of estimated and observed values.

Pearson correlation coefficient (r)

Pearson's coefficient of correlation was discovered by Bravais in 1846, but Karl Pearson was the first to describe, in 1896, the standard method of its calculation and showed that it was the best one possible. An important assumption in Pearson's 1896 contribution is the normality of the variables analyzed, which could be true only for quantitative variables. Pearson's correlation coefficient is a measure of the strength of the linear relationship between two such variables [10]. The value of correlation coefficient ranges between -1 and 1. If the two variables are in perfect linear relationship, the correlation coefficient will be either 1 or -1. The sign depends on whether the variables are positively or negatively related. The correlation coefficient is 0 if there is no linear relationship between the variables. Given paired measurements (x_1,y_1), (x_2,y_2). . . (x_n,y_n), the Pearson correlation coefficient is given by

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \qquad - Eq(1)$$

Where $\bar{x}$ and $\bar{y}$ are the sample mean of x1,x2….xn and y1, y2…..yn, respectively.

Recession method

In the absence of recharge, river flow is supported by groundwater contribution often termed baseflow. In some parts of the world river flow is seasonal and baseflow in rivers is an important water resource; the characteristics of which must be fully understood to ensure an optimum management of water supply during prolonged droughts.

During periods of recession, the flow exhibits a pattern of exponential decay giving a curved trace on a simple plot of Discharge versus Time, t. The equations may take different forms but the most commonly used are:

The slope of flow recession, $\alpha = \dfrac{\ln Q_{t_0} - \ln Q_{t_1}}{t_1 - t_0}$ $\qquad - Eq(2)$

The base flow recession constant, $k = \dfrac{1}{\alpha}$ $\qquad - Eq(3)$

Hence at time, t within the gap,

$$Q = {}^{\circ}Q_{t_0} {}^{\circ}\exp\left(-\frac{t - t_0}{k}\right) \qquad - Eq(4)$$

To use equation (4) as a predictive model, the base flow recession constant, k has to be determined for the catchment. This is done by plotting the exponential graph of the dry weather discharges against time, and then calculates k from the slope of flow recession. The accuracy of Qt thus depends among other things on the accuracy of k. The size of die time units, t (i.e., whether hourly, daily, weekly, monthly, etc) also affects the accuracy of the predicted flow [11,12].

Regression method

For many rivers, downstream or upstream data are missing. In this case the flow data from the nearby rivers can be used to estimate the missing flows. The regression analysis is the method frequently used to solve this problem. The dependent variables are the flows of the nearby stations having drainage basins with similar hydrological characteristics [13].

These methods for analyzing multivariate data with missing observations from one or more independent variables are presented by Beale et al. and Johnson et al. [14,15].

Linear regression analysis: If X and Y are two related variables, then linear regression analysis helps us to predict the value of Y for a given value of X or vice versa, expressed by

$y = \beta x + \alpha$ $\qquad$ -Eq(5)

Where by β is the slope and α is y-intercept

Multiple regression analysis: This helps to predict the value of y for given values of x_1, x_2, x_3…x_k.

Expressed by

$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \ldots\ldots\ldots \beta_k x_k + \alpha$ $\qquad$ -Eq(6)

Whereby; β_1 β_2 β_3 β_k are parameters to be determined and α is y-intercept.

Nash-Sutcliffe efficiency

The Performances of selected method were assessed using Nash-Sutcliffe model efficiency coefficient (NSE). It is commonly used to assess the predictive power of hydrological discharge models. It is defined as:

$$E = 1 - \frac{\sum_{i=1}^{n}(x_{obs,i} - x_{estm})^2}{\sum_{i=1}^{n}(x_{obs,i} - \bar{x}_{obs})^2} \qquad - Eq(7)$$

The value of E, ranges from -∞ to 1. Where xobs is observed values and xestm is estimated values at time/place i and $\bar{x}_{obs}$ is the average value of observations. Essentially, the closer the model efficiency is to 1, the more accurate the model.

Root mean square error (RMSE) and mean absolute error (MAE)

The root mean square error (RMSE) has been used as a standard statistical metric to measure model performance in meteorology, air quality, and climate research studies. The mean absolute error (MAE) is another useful measure widely used in model evaluations [16]. RMSE presents information on the short-term efficiency which is a benchmark of the difference of predicated values about the observed values. The lower the RMSE, the more accurate is the evaluation. MAE (mean absolute error) is an indication of the average deviation of the

predicted values from the corresponding observed values and can present information on long term performance of the models; the lower MAE the better is the long-term model prediction [17].

While the MAE gives the same weight to all errors, the RMSE penalizes variance as it gives errors with larger absolute values more weight than errors with smaller absolute values. When both metrics are calculated, the RMSE is by definition never smaller than the MAE [16]. The RMSE and MAE are given by the following formula;

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_{Obs,i} - X_{estm,i})^2}{n}} \quad - Eq(8)$$

Where $x_{obs,i}$ is observed values and $x_{estm,i}$ is estimated values at time/place i.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right| \quad - Eq(9)$$

Results and Discussion

Selection of dependent and independent variables

For studied gauging stations, the correlation coefficients were determined by equation (1). These ranged from 0.4 to 0.9 indicating fairly good relationship between gauging stations as shown in Table 2. With regard to gauging stations in Figure 1, generally nearby gauging stations or stations in the same river experience the same river flow pattern (Figure 2a). These indicated strong correlation compared to stations which are far apart (Figure 2b). The Gauging stations with strong correlation estimate the missing river flow data for each other more efficiently compared to those with poor correlation. This has been supported by DeSiva et al. in a Comparison of Methods Used in Estimating Missing Rainfall Data [6].

Gauging Stations	Mtitu	Salimwani	Kimani	Ndiuka	Mtandika	Igawa	Ilongo	Ihumbu	Makalala	Msembe	Chitekero
Chitekero (1KA7A)	0.700	0.721	0.613	0.696	0.566	0.696	0.738	0.728	0.703	0.566	-
Msembe (1KA59)	0.481	0.534	0.440	0.565	0.390	0.523	0.593	0.490	0.526	-	0.566
Makalala (1KA32A)	0.735	0.535	0.610	0.815	0.579	0.772	0.807	0.902	-	0.526	0.703
Ihumbu (1KA21A)	0.792	0.579	0.612	0.799	0.581	0.760	0.816	-	0.902	0.490	0.728
Ilongo (1KA15A)	0.710	0.631	0.655	0.720	0.520	0.765	-	0.816	0.807	0.593	0.738
Igawa (1KA11A)	0.705	0.606	0.727	0.745	0.489	-	0.765	0.760	0.772	0.523	0.696
Mtandika (1KA37A)	0.694	0.426	0.371	0.630	-	0.489	0.520	0.581	0.579	0.390	0.566
Ndiuka (1KA2A)	0.753	0.473	0.658	-	0.630	0.745	0.720	0.760	0.815	0.565	0.696
Kimani (1KA9)	0.583	0.538	-	0.655	0.371	0.727	0.655	0.612	0.610	0.439	0.613
Salimwani(1KA8A)	0.577	-	0.538	0.473	0.426	0.606	0.631	0.579	0.535	0.534	0.721
Mtitu (1KA22)	-	0.577	0.583	0.753	0.694	0.705	0.710	0.728	0.735	0.481	0.700

Table 2: Correlation coefficients between gauging stations.

Figure 2a: A graph showing river flow pattern between gauging stations with strong correlation.

Figure 2b: A graph showing river flow pattern between gauging stations with weak correlation.

Comparison of estimates

The estimates obtained from each method are compared with observed records. The suitability of method is decided by how close the estimates and observed values are in a given time series. Several descriptive statistics of error as indicated in Tables 3-5 are used as criteria to evaluate the efficiency of the method in estimating missing data.

Linear and multiple regression methods: Different techniques offer different performances, according to the characteristics of initial data points. Linear and Multiple regression have been used widely in estimating missing data during period of river flow fluctuation. The value for NSE, RMSE and MAE were determined using equation (5) through equation (9). The results are presented in Tables 3 and 4 for Linear and Multiple regression methods respectively.

| Independent variable	Dependent variable	NSE	RMSE	MAE
Ilongo (1KA15A)	Chitekero (IKA7A)	0.64	1.69	0.95
Ilongo (1KA15A)	Msembe (1KA59)	0.42	53.97	28.55
Ihimbu (1KA21A)	Makalala (1KA32A)	0.81	1.21	0.77
Makalala (1KA32A)	Ihimbu (1KA21A)	0.81	5.14	2.76
Mtitu(1KA22)	Mtandika (1KA37A)	0.60	6.69	3.97
Makalala (1KA32A)	Ilongo (1KA15A)	0.76	2.72	1.78
Ilongo (1KA15A)	Igawa (1KA11A)	0.64	8.14	4.20
Makalala (1KA32A)	Ndiuka (1KA2A)	0.56	8.87	7.03
Igawa (1KA11A)	Kimani (1KA9)	0.47	9.75	4.00
Ihimbu (1KA21A)	Salimwani(1KA8A)	0.57	26.40	17.23

Table 3: Results of NSE, RMSE and MAE for Linear Regression Method.

| Independent variable	Dependent variable	NSE	RMSE	MAE
Ilongo (1KA15A)	Chitekero (IKA7A)	0.65	1.68	0.91
Ihimbu (1KA21A)				
Ihimbu (1KA21A)	Msembe (1KA59)	0.46	56.90	31.31
Makalala (1KA32A)				
Mtitu(1KA22)	Makalala (1KA32A)	0.81	1.20	0.76
Ihimbu (1KA21A)				
Makalala (1KA32A)	Ihimbu (1KA21A)	0.88	4.00	2.02
Mtitu(1KA22)				
Chitekero (IKA7A)	Mtandika (1KA37A)	0.63	6.48	3.91
Mtitu(1KA22)				
Ihimbu (1KA21A)	Ilongo (1KA15A)	0.78	2.59	1.85
Makalala (1KA32A)				
Ilongo (1KA15A)	Igawa (1KA11A)	0.71	7.28	3.34
Ihimbu (1KA21A)				
Makalala (1KA32A)	Ndiuka (1KA2A)	0.66	7.74	6.24
Ihimbu (1KA21A)				
Ilongo (1KA15A)	Kimani (1KA9A)	0.47	9.75	3.97
Igawa (1KA11A)				

Ilongo (1KA15A)	Salimwani(1KA8A)	0.68	21.90	10.32
Ihimbu (1KA21A)				

Table 4: Results of NSE, RMSE and MAE for Multiple Regression Method.

From Figure 3, 1KA32A and 1KA22 have been used to estimate missing data for station 1KA21A. The results show that Multiple regression (Figure 3a) has the better estimate with NSE (0.88), RMSE (4.003) and MAE (2.023) compared to Linear regression (Figure 3b) by using 1KA32A which gave the NSE (0.805), RMSE (5.137) and MAE (2.758). On the other hand, there is no significant difference between NSE value for Linear and Multiple Regression Analysis this gives the evidence that in absence of more than one independent variable, Linear Regression can be employed to estimate the missing data of river flow.
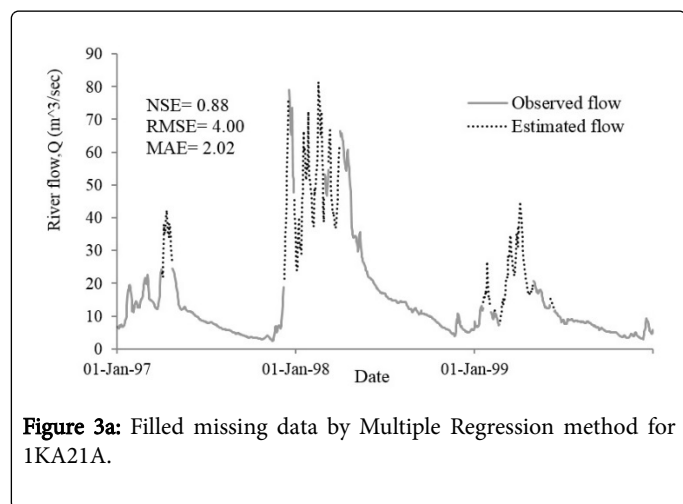

**Figure 3a:** Filled missing data by Multiple Regression method for 1KA21A.
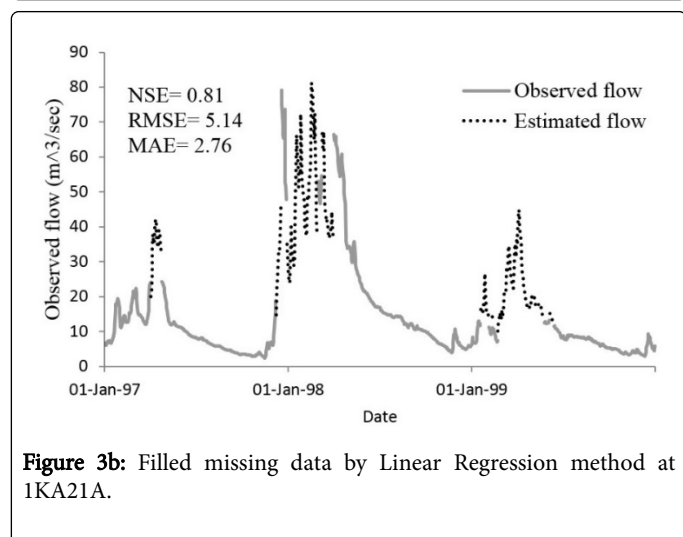
Figure 3b: Filled missing data by Linear Regression method at 1KA21A.

Also, a good estimator for one station is not necessary to be a good estimator for another station. From the Table 3, 1KA32A when used to estimate missing data for 1KA21A the efficient of estimation (NSE) was 0.805, the same independent variable (1KA32A) when used to estimate the missing data for 1KA2A the efficient of estimation (NSE) was 0.558. The results are presented graphically on Figures 4a and 4b.
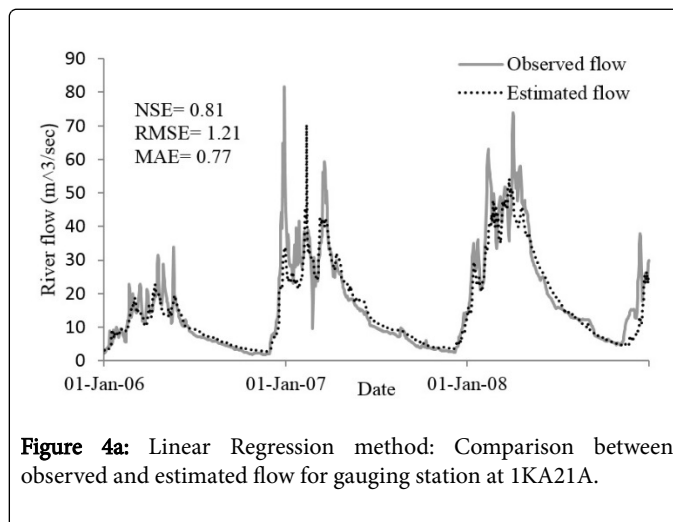
Figure 4a: Linear Regression method: Comparison between observed and estimated flow for gauging station at 1KA21A.
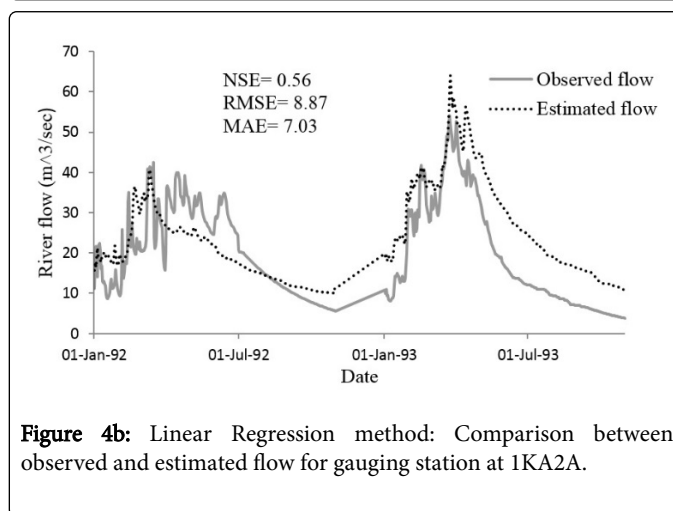
Figure 4b: Linear Regression method: Comparison between observed and estimated flow for gauging station at 1KA2A.

In addition, it has been found that stations in the same river can estimate each other better than choosing station from different rivers. From Figure 5a, 1KA32A and 1KA22 in the same river (Little Ruaha River) were used to estimate missing data for 1KA21A and NSE (0.88) was obtained. In contrary from Figure 5b, the gauging station 1KA15A (River Ndembera) and 1KA11A (River Mbarali) from different rivers were used to estimate the missing data for 1KA9A (River Kimani) and the estimate was poor with NSE of 0.47.
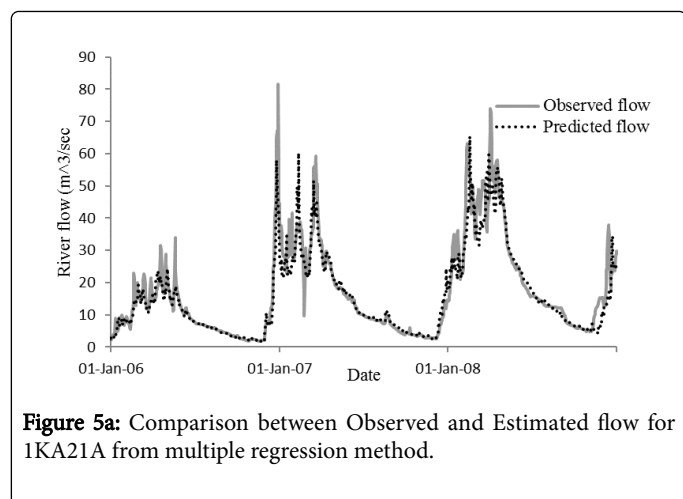
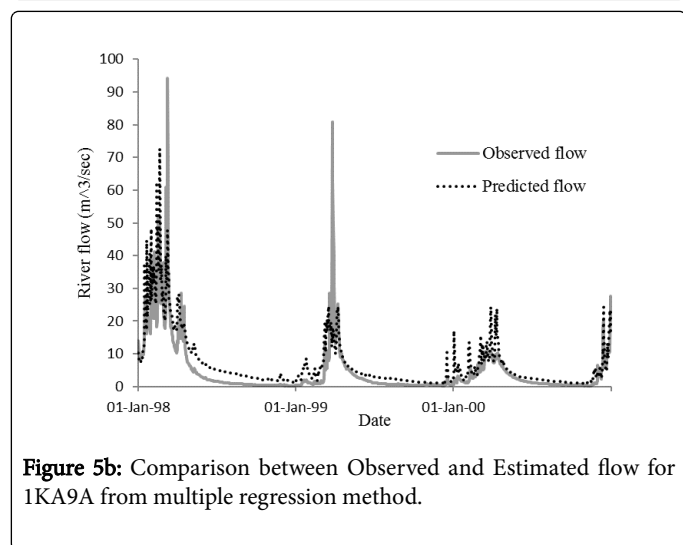**Figure 5a:** Comparison between Observed and Estimated flow for 1KA21A from multiple regression method.

Figure 5b: Comparison between Observed and Estimated flow for 1KA9A from multiple regression method.

Recession method: From the study, Recession method has proven to be the best in filling missing data during period of low flow. The values for recession factors, are shown in the Table 5. The best estimate for recession method was found on Makalala (1KA32A) gauging station (Figure 6a) while the poor estimate was found on Kimani (1KA9A) gauging station (Figure 6b).

The poor estimate at Kimani (1KA9A) gauging station is due to variability in river flow pattern for the five years period of calibration that leads to great variation of recession factors of river flow during the dry period for each year.
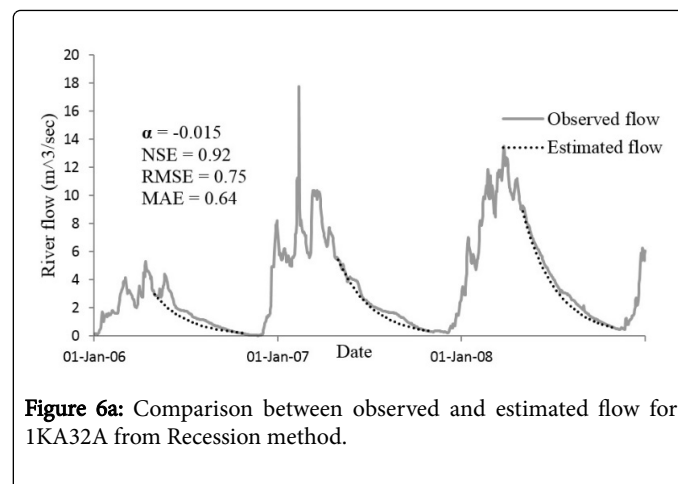

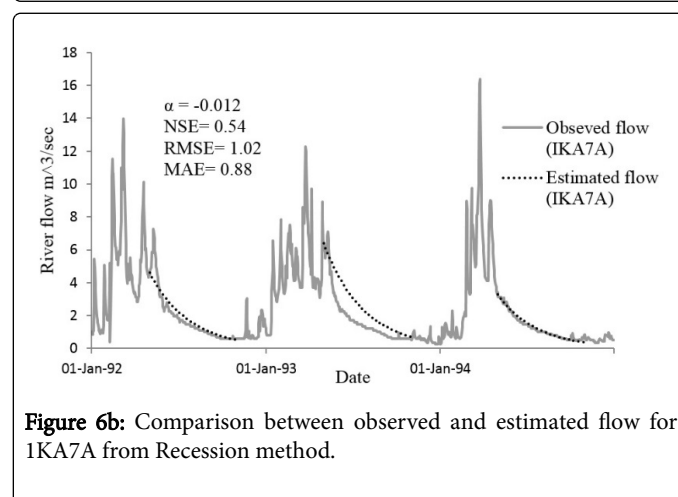**Figure 6a:** Comparison between observed and estimated flow for 1KA32A from Recession method.

Figure 6b: Comparison between observed and estimated flow for 1KA7A from Recession method.

Station	Recession factor, α	NSE	RMSE	MAE
Chitekero (IKA7A)	-0.012	0.54	1.02	0.88
Msembe (1KA59)	-0.040	0.85	11.02	7.35
Makalala (1KA32A)	-0.015	0.92	0.75	0.64
Ihimbu (1KA21A)	-0.009	0.88	1.39	1.04
Mtandika (1KA37A)	-0.005	0.88	1.90	1.26
Ilongo (1KA15A)	-0.020	0.82	1.29	1.17
Igawa (1KA11A)	-0.009	0.83	1.76	1.50
Ndiuka (1KA2A)	-0.011	0.64	3.73	3.24

Kimani(1KA9A)	-0.014	0.75	0.74	0.62
Salimwani (IKA8A)	-0.009	0.65	1.14	1.00

Table 5: Results of Recession factor (α), NSE, RMSE and MAE for Recession Method.

Conclusion

The choice of the method to be utilized to fill missing river flow data depends on the data availability between the variables. In this study, the estimates were determined by correlation coefficient between the stations, as well the distances between the stations where the nearer stations were able to estimate each other better than the far ones. Also, the stations which are located in the same stream were able to estimate each other better than stations which belong to different streams.

In filling the missing river flow data, a single gauging station can be filled with more than one method. Multiple regressions are the best method in filling missing data during period of high flow compared to Linear regression method, however in the area where the correlated stations are scares linear regression can be employed.

References

1. Mwaruvanda W (2009) The great ruaha river profile. A Paper presented at the Clivet Project Inception Workshop, Blue Pearl Hotel Ubungo Plaza, Dar es Salaam, 27th November.
2. Kashaigili K, Japhet J, Kadigi RMJ, Lankford BA, Mahoo HF, et al. (2005) Environmental Flows Allocation in River Basins: Exploring Allocation Challenges and Options in the Great Ruaha River Catchment in Tanzania. Physics and Chemistry of the Earth 30: 689-697.
3. Elshorbagy AA, Panu US, Simonovic SP (2000) Group-based estimation of missing hydrological data: I. Approach and general methodology. Hydrological Sciences Journal 45: 849-866.
4. Presti RL, Barca E, Passarella G (2010) A methodology for treating missing data applied to dailyrainfall data in the Candelaro River Basin (Italy). Environmental Monitoring and Assessment 160: 1-22.
5. Hasana M, Crokea B (2013) Filling gaps in daily rainfall data: a statistical approach. Mssanz Org Au, pp: 1-6.
6. DeSilva RP, Dayawansa NDK, Ratnasiri MD (2007) A comparison of methods used in estimating missing rainfall data. Journal of Agricultural Science 3: 101-108.
7. Mtahiko MGG, Gereta E, Kajuni AR, Chiombola EAT (2006) Towards an Ecohydrology-Based Restoration of the Usangu Wetlands and the Great Ruaha River, Tanzania. Wetlands Ecology and Management.
8. Willmott CJ, Matsuura K (2005) Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research 30: 79-82.
9. Elshorbagy AA, Panu US, Simonovic SP (2000) Group-based estimation of missing hydrological data: II. Application to streamflows. Hydrological Sciences Journal 45: 867-880.
10. Hauke J, Kossowski T (2011) Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data. Quaestiones Geographicae 30: 87-93.
11. Oyegoke ES, Ige ED, Oyebande L (1983) River flow forecasting through a regression model: a case study of the basement complex of western Nigeria. Proc. Hamburg Symposium, IAHS Publication 140: 419-429.
12. Bako MD, Hunt DN (1988) Derivation of Baseflow Recession Constant Using Computer and Numerical Analysis. Hydrological Sciences Journal 33: 357-367.
13. Cigizoglu HK (2003) Estimation, forecasting and extrapolation of river flows by artificial neural networks. Hydrological Sciences Journal 48: 349-362.
14. Beale EML, Little RJA (1975) Missing values in multivariate analysis. J Roy Statist Soc B 37: 129-145.
15. Johnson RA, Wichern DW (1988) Applied Multivariate Satistical Analysis. Prentice Hall, New Jersey, USA.
16. Chai T, Draxler RR (2014) Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. Geoscientific Model Development 7: 1247-1250.
17. Doreswamy D, Vastrad CM (2013) Performance Analysis of Neural Network Models for Oxazolines and oxazoles derivatives descriptor dataset. International Journal of Information and Techniques 3: 1-6.