Addis Ababa University: Institute of Technology

School of Graduate Studies

School of Civil and Environmental Engineering

Research Proposal on

Low Flow Analysis for Ungauged catchment in Baro Akobo river basin, Ethiopia

Submitted to: Addis Ababa University-Horn of Africa Regional Environment Centre and Network (HoA-REC&N)

By: Kidist Assefa

Advisor: Belete Berhanu (PhD)

Submitted on: Agust, 2021

Addis Ababa – Ethiopia

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1. Background of the Study

Low-flows are an important part of the natural flow regime of rivers and streams. During water resource planning and design the spatial and temporal variability of river low flow characteristics should be considered. A high economic value is associated with the activities of prediction and analysis low flow and the resulting long-term drought. It should be mentioned that drought have more severe consequences and are often costlier than the flood events (Sustainet E.A., 2010). Decreasing of an environmental low flow in a given ecosystem can affect multi-purpose operations which depends on that water system either large river or lakes. Some of the impacts include severe consequences in water quality and river ecological status and Navigation and power supply sectors (M. J. Kirkby et al, 2011).

Stream flow and components (Such as low flows) information is required for the assessment of the impact development on many water resource applications (S. G. Gebrehiwot et al, 2011). Information for an ungagged site may be obtained either by means of deterministic or Stochastics models by regionalization techniques based on available observed flow records from gauged stations. In the first case a satisfactorily long daily flow time sequence is simulated and then any required flow characteristic is calculated from the simulated series (E. Toth, 2013).

1.2. Problem Statement

Evaluation of low flow is very vital for water resource development and management for sustainable and proper utilization of water resource. Low flow is one of the essential and key parameters often uses to plan several water resource developments such as drinking-water supplies, irrigation, to generate electricity, ecosystem restoration, recreation and transportation beyond home for several aquatic living things. Majority of the water resource development activities require low flow data to design the infrastructure and to quantify the water supply and demand through characterizing the quantity, timing, and frequency depending on the location of interest. It is difficult to get since most of the river systems are un-gauged or poorly gauged catchments. Although, to estimate the low flow a good quality of data for a long period is one of the most deterrent factors especially in developing countries since most of the rivers are un-gauged or poorly gauged catchments. Therefore, quantifying the low flow for the ungauged catchment of a given river basin is crucial for biophysical modeling and to estimate different components of the water cycle. Baro Akobo River basin is located in the southwestern part of Ethiopia, and it has huge potential for fresh water accessibility. However, there are limited stream flow gauging stations and most of its subbasins are ungauged from development point of view. Therefore, this study will develop the low flow estimation method for ungauged catchments in Baro-Akobo River in Ethiopian using mainly the remote sensing and rainfall-runoff modeling approach and establish virtual gauging stations.

1.3. Objectives

The main objective of this study is to develop reliable and easily applicable method or approach to estimate low flows of ungauged rivers or streams in Baro Akobo basin.

1.3.1 The specific objective

- To estimate daily streamflow for ungauged river using the-state-of-the-art rainfall runoff modeling and remote sensing approach
- To develop low flow separation method appropriate for ungauged river system.
- To establish low flow regimes classification of the Baro Akobo basin.

1.4. Significance of the study

Low flows are important indicators of river flow/stream health in a given ecosystem. It is also important for water resource development and optimize the water resources for sustainable utilization of the resource. In the Baro Akobo basin estimation of the low flow is important as the water sector development and use are progressing. In order to optimize and asses' status of low flows evaluating the regimes and regionalizing low flows for the ungagged part of the Ethiopia is very crucial in data scarce. The maximum water resource potential of the Baro Akobo is for over short seasons which is three months and this water does not stay for longer in some part of the basin. This situation requires utilization dry season flow for Agriculture to improve livelihoods through irrigation. Even though there are studies in the Baro Akobo basin focus on the maximum water contents and no research has been conducted for this living. But shortage amount of water periods and this research will have used to indicate flow analysis scenario of the Baro Akobo basin and give remain about the low flow water resource management.

2. LITERATURE REVIEW

2.1. Low flow

The continuously varying flow of a stream is characterized by the mean and variability of daily, monthly, and annual flows. Stream flow records are published as daily mean flows from which the monthly means and the annual mean are computed. The mean monthly flow for all twelve months shows the flow distribution within the year. Stream flow at a stream gauge could be affected by many natural factors including climate, geomorphology of the channel, geology, and groundwater levels. Human activities such as building reservoirs and dams, and making withdrawals from surface-water and groundwater systems for municipal and industrial use also affects stream flow (WMO., 2009a)

Usually the low-flow series extraction models include 3d-slf, 7 d-slf and 14d-slf. Many local and international water authorities in the world use a frequency statistic computed by using an annual 7 consecutive-day series to assist them in making regulatory decisions. Flow frequency analysis of the annual 7-day low flows provides a means for estimating the probability of occurrence of a given discharge for each climatic year of data available. Low-flow frequency commonly is referred in terms of recurrence interval or the probability of being exceeded.

2.2. Importance of Low Flow

Population growth and the associated expansion in domestic, industrial and agricultural use of water has placed an increasing demand on water resources throughout the world. More recently, greater awareness of the ecological, recreational and amenity benefits of the river corridor and concerns over changes in river regimes in the longer term have added to the pressures on water resources. It is in times of droughts, or in other words in low flow periods, that river systems are most stressed and thus an understanding of the natural variability of drought conditions in **time** and **space** is fundamental to a wide range of water management problems.

In the early days of application of analytical techniques in the field of hydrology, low flow studies of a river/stream were basically carried out to analyses its ability to supply a particular water demand. The main categories of water demand were, (i) Domestic, (ii) Agricultural and (iii)

Industrial. If it was found that the flow of the river was insufficient to supply a demand, further analysis was carried out to ascertain how much water needed to be stored in order to meet the demand. Even in the present context, low flow studies remain important for the supplying of such direct demands, either directly or by a suitable storage. However, the greater awareness of environmental impacts of the use or the abuse of water courses by mankind such as the disposal of industrial and domestic effluents has oriented low flow studies in another important direction. Extreme low flow events are more diligently analyzed these days in the context of ascertaining whether a certain water course can take up a given load of effluents. The effect of very low flows on the ecological and recreational aspects of rivers also adds to the newly emerging importance of low flow studies.

2.3. Low Flow Analysis

Low flow analysis, the determination of minimum flows at different recurrence interval, is a common problem in hydrology. The standard procedure to determine probabilities of low flows consists of fitting the observed stream flow record to specific probability distributions. Low flow analysis methodologies include collection of important data for the study area, Data quality checking for the selected stations.

Daily flow information is required to assess the impact of development on water resource. Information for ungauged site may be obtained either by means of deterministic models or by regionalization techniques based on available observed flow records from gauged stations. In addition, the frequency and estimation of low flows has also been hardly studied in the basin. Therefore, this study will have two main focuses including i) low flow separation for the ungauged catchments of the basin and establish virtual gauging stations ii) carrying out low flow estimation.

2.4. Base flow separation

Base flow separation uses the time-series record of stream flow to derive the base flow signature. Graphical separation methods tend to focus on defining the points where base flow intersects the rising and falling limbs of the quick flow response. Filtering methods process the entire stream hydrograph to derive a base flow hydrograph. Recursive digital filters, which are routine tools in signal analysis, are commonly used to remove the high-frequency quick flow signal to derive a low-frequency base flow signal. Such filters are simple and robust but the results are very sensitive to the filter parameter, which needs calibration before the results can be considered to be numerically valid.

As the hydrographic record represents a net water balance, base flow is also influenced by any water losses from the stream such as direct evaporation, transpiration from riparian vegetation, or seepage into aquifers along specific reaches. Water use or management activities such as stream regulation, direct water extraction, or nearby groundwater pumping can significantly alter the base flow component. Hence, careful consideration of the overall water budget and management regime for the stream is required (Hakan T. and Martijn J.B., 2018).

The direct flow is primarily the direct response of a rainfall event and includes the overland flow (runoff) and lateral flow in the soil profile also known as interflow. The base flow is a component of stream flow which is discharged from natural storage aquifers. A stream flow can be affected by the interruption of the direct flow, such as by diversion of runoff and water harvesting mechanisms. Moreover, in order to have targeted policy in water resources development and use, it is essential to have a reasonable estimation of the direct and base flow components of stream flow (Ruigang Z. et al., 2012).

2.5. Model Selection

The selection of a particular model is a key issue to get satisfactory answers to a given problem. Currently, there are numerous hydrological models simulating the hydrological process at different spatial and temporal scales. Although there are no clear rules for making a choice between models, some simple guidelines can be stated. Starting from the studied physical system, the first step is to define the problem and determine what information and questions need to be answered. This means it is necessary to evaluate the required output, the hydrologic processes that need to be modelled, availability of input data. Subsequently the simplest method that can provide the answer to the questions has to be chosen (Singh, V.P. and Woolhiser D.A, 2002). Generally, the choice of the best model depends to a large extent on the problem. The nature of the physical processes involved, the use to be made of the model, The quality of the data available and The decisions that rest on the outcome of the model's use.

a) GR4J (Ge`nie Rular a` Parame`ter Journalier)

GR4J (Ge´nie Rural `a 4 param`etres Journalier) is rain fall runoff hydrological model which can be used as tools to better understand the behavior of different components of the hydrological cycle (L. Coron et al., 2017). One of GR4J's unique aspects is the use of a "groundwater exchange coefficient". When this parameter is away from zero, the model relaxes the requirement that runoff balances rain fall and evapotranspiration. Although the parameter is conceptually linked to groundwater, showed that it is also effective at reducing model bias, more so than simple scaling parameters for rainfall, potential evapotranspiration or runoff (Crochemore L., 2016). The model contains two stores and has four parameters' inputs are daily rainfall is an estimate of areal rainfall, e.g., calculated each day as an average on all the available rain gauges. The model simulates daily flow time series. Potential evapotranspiration (PE) time series PE can be a long-term average regime length, i.e., the same PE length is used every year. Observed daily flow time series are required for model calibration and evaluation. Overall model has four free parameters: X_1 maximum capacity of the production store (mm), X_2 groundwater exchange coefficient (mm), X_3 one-day-ahead maximum capacity of the routing store (mm), X_4 time base of unit hydrograph UH1 (days) (Perrin, C. et al, 2015).

2.6. Virtual Measurement

The Virtual Measurement program are metrology constructs standard reference computation, uncertainty quantification and traceability into scientific computation and computer assisted measurement technologies. As with physical measurement systems, development of a virtual metrology infrastructure will result in predictive computing with quantified reliability. In turn, this will enable improved decision making based on results of computer simulations.

2.6.1 Application of Remote sensing for estimating of streamflow data

I. Estimation of Streamflow and Classifications

Streamflow in a given catchment is the aggregated result of all geological and climatological factors that operate in that catchment. The availability of streamflow data is very important for planning, designing and operating of any water resource development and management project. Therefore, measuring streamflow is important to good water management practices provides higher confidence than measurement of other components (Verma, R. K. et al, 2017).

Stream gauges have been used to measure the amount of streamflow at a particular location of a river or a stream at regular time intervals. Despite the importance of streamflow data, the total number of stream gauges across many parts of the world is declining. To complicate matters further, existing gauges are at the outset sparse and poor management practices that either streamflow data are not recorded at all or false information is recorded. The false recording of streamflow data in some countries has more adverse effects than the non-availability of data and has serious effects on the planning, designing and evaluation of water resource projects (N.T.S. Wijesekera, 2018). This situation is exacerbated due to increased negligence in maintaining the streamflow gauges properly especially since the benefits of such gauging network are invisible and difficult to account for accurately (K.L. Jaeger et al., 2018).

II. Remote sensing

Streamflow is a fundamental quantity to evaluate hydrological processes and is a quantitative indicator for integrated water management systems. Measurement of river discharge is necessary in hydrology and water resource management. The knowledge of river flow propagation speed and the time for flows to pass downstream is critical for streamflow forecasting, reservoir operation, hydropower, irrigation, water supply and watershed modeling. Therefore, there is a great need for long-term, continuous, spatially consistent, and readily available discharge data. In the absence of such measured and estimation streamflow data using meteorological data represents a viable alternative. Nevertheless, this alternative is not always possible due to the unavailability of required meteorological data. In the face of such data limitations, many have advocated the use of remote sensing (RS) data to estimate streamflow (Arthur W. S et al., 2018).

Unlike ground observations, remote sensing provides a limitless coverage of rivers and other water bodies. As evidenced by recent studies, remote sensing has a good spatial coverage and a long monitoring period, both of which have contributed to a growing interest in deriving discharge estimates from remote sensing. The ground observation method is the most accurate measure of river discharge. However, ground river discharge is obtained by estimating the hydraulic characteristics of stream channels including depth, width, and velocity (Velpuri N. M. et al, 2012).

2.7. Singular spectrum analysis

Singular spectrum analysis (SSA) is a technique of time series analysis and forecasting. It combines elements of classical analysis, multivariate statistics and geometry, dynamical systems and signal processing. From the algorithmic point of view, SSA can be considered as a typical subspace-based method of signal processing (Hassani, H., et al, 2014).

SSA important than other methods First the modelling procedure employed by SSA and classical time series techniques. The classical time series methods consider modelling and forecasting. However, the SSA technique will filter data such that the trend, signal and noise could be identified separately. Thereafter, SSA reconstructs a new time series which corresponds to a less noisy approximation of the signal, for generating forecasts (Hassani, H., et al., 2013a). Secondly, SSA is a nonparametric technique which does not rely on the assumptions of normality for the residuals, stationarity of the data, and linearity for the model which are highly unlikely to hold in real world applications (Silva, E. S. and Hassani, H., 2015).

2.8. Regionalization Low flow analysis

Regionalization is the process of transferring information from gauged catchment to ungauged catchment. Estimation of stream flow in ungauged basins commonly is based on the principles of regionalization (Gregor L., 2014). Prediction curves can be established at ungauged sites by standardizing the flow at gauged sites by a scale index and by combining the information into a single regional curve. It is also possible to synthetically generate a large number of single or multi-site streamflow time series data based on other gauging station records and then proceed with the estimation of the low flow characteristics (Ni. L.W. and Khin M.W., 2014).

The availability of data is an important aspect in frequency analysis. In practice, however, data may be limited or in some cases may not be available for a site. In such cases, regional analysis is most useful. The two probably most popular approaches are based on principles of similarity by spatial proximity and on similarity of catchment characteristics. The first approach is based on the rationale that catchments of close proximity will likely have a similar flow regime since climatic, topographic, and physiographic settings are comparable. The second approach is based on the assumption that optimized parameters representing certain catchment characteristics are also applicable in other catchments with similar characteristics (Sang Ug K. and Kil S., 2010).

2.8.1 Fuzzy clustering algorithm

The fuzzy c-means algorithm is very similar to the k-means technique except that it computes the degree to which a site belongs to each cluster on a scale of 0 to 1 where a value of 1 represents full membership. As such, each climate station can partially belong to several clusters theoretically providing a more accurate partitioning of the sites. The outcomes of the regionalization procedure often require subjective and manual adjustments to site membership in order to improve the regional of flow variability to an acceptable level. Its membership function provides useful information for removing or relocating discordant stations; thereby offering another advantage of the fuzzy c-means algorithm over traditional hard clustering techniques (Srinivas, 2013). It Conducted comparative analyses between the fuzzy c-means and k-means algorithms for flow regimes. They concluded that the former technique outperformed the latter by achieving a greater number of homogeous hydrologic regions. Although the analyses were conducted for the regionalization their findings also apply to the formation of flow regions.

3. MATERIALS AND METHODS

3.1. Description of the study area

The Baro Akobo basin is located in the southwestern part of Ethiopia. it is located, between latitudes 5° 31" and 10° 54" north and longitude 33° 0" and 36° 17"east and covers 76,000 km^2. This area includes all or part of the four administrative regions: SNNPRS (Southern Nations & Nationalities People Regional State) in the south, Oromiya in the northeast, Gambela in the central western part and Benishangul Gumuz in the northwestern extremity. Based on the basin bordered is the South Sudan in the West, northwest and southwest, Abbay and Omo-Ghibe Basins in the east.

The Baro-Akobo basin is the fourth largest basin in the country, next to the Wabi Shebelle, Abbay and Tekeze river basins. One of the tributaries of the Baro River is a river in southwestern Ethiopia, which defines part of Ethiopia's border with South Sudan. Elevation of the study area varies between 440 and 3000 m a.m.s.l. The higher elevation ranges are located in the North East and Eastern part of the basin while the remaining part of the basin is found in lower elevation. It is high variability in temperature with large differences between the daily maximum and minimum temperatures. The river's mean annual discharge at its mouth is 241 m³/s.

Figure 1: Location of the study area

3.2. Materials and Methodology

The general methodology used for this study includes three main procedures after selected the stations and checking the data quality as: The first separation of flow data, second estimation of low flow and third regionalization and classification of ungauged flow regime.

3.2.1 Data Availability and Sources

Most of catchments in Ethiopia are data scarce, but one of the selection criteria is the catchments which have some ground measured data. Various types of data collected from different organization in order to use for low flow analysis. The main input data needed for this study involves temporal and spatial data. The temporal data consists of the time series Hydrological and Meteorological data. The spatial data mainly consists of digital elevation model (DEM), Soil map, and land use map.

Table 1: Data type and source for the study area

S/No.	Data type	Source	Purpose
1	Temporal data	MoWIE and National Metrological Agency	For analysis
2	Spatial data	MoWIE	For analysis
3	Remote Sensing	Planet: Planet Explorer https://earthengine.google.com https://glovis.usgs.gov/app and ALASKA	For analysis

3.2.2 Data Quality Cheek

Spatial Data Collection and Analysis

The three types of spatial data namely Digital Elevation Model (DEM), soil and land use/ land cover map were collected from GIS department of MoWIE.

Digital Elevations Model

Topography is defined by a Digital Elevation Model (DEM) that describes the elevation of any point in a given area at a specific spatial resolution. It is also widely known as a Digital Terrain Model (DTM). A DEM can be represented as a raster (a grid of squares) or as a triangular irregular network. DEMs are used often in geographic information systems, and are the most common basis for digitally-produced relief maps. In this study DEM which is 12.5 by 12.5-meter resolution was downloaded from **ALOS PALSAR** data site. The DEM were used to delineate the watershed, to extract information about the topography/elevation of the watershed and to analyze the drainage patterns of the land surface terrain. Sub-basin parameters such as slope gradient, slope length of the terrain, and the stream network characteristics were derived from the DEM. Some flow data collected from Ministry of water Resource of hydrology department and the remaining Parts were digitized during the study time.

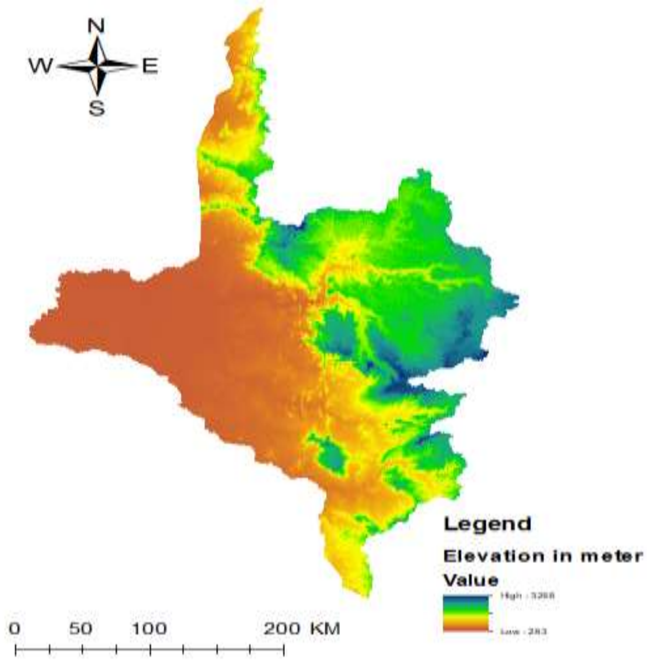

Figure 2: Topographic Variation

Soil

Soil data is the other imputes required for hydrological and statistical modeling which influences the runoff generation of catchment. The physical property of the soil in each horizon governs the movement of water and air through the soil profile, daily runoff and erosion. It has a major impact on cycling of water in hydrologic response unit (HRU), and used to determine water budget for the soil profile, daily runoff and erosion.

Land use/land cover

Land use/ Land cover is one of the most important factors that affect runoff, evapo-transpiration and surface erosion in a watershed. Land cover data documents how much of a region is covered by forests, wetlands, impervious surfaces, agriculture, and other land and water types. Land use and land cover have some fundamental differences. Land use refers to the purpose the land serves, for example, recreation, wildlife habitat or agriculture; it does not describe the surface cover on the ground. For example, a recreational land use could occur in a forest, shrub land, grasslands or on manicured lawns. Land use and cover condition can affect the hydrological balance of the watershed by changing magnitude and pattern of runoff, peak flow and ground water levels.

Temporal Data Collection and Analysis

Engineering studies of water resource development and management using hydrological model depends critically on hydrological and meteorological data (model calibration more than anything relies on these data quality). Therefore, estimating of missing, checking the availability, consistency, and homogeneity of data were executed to enhance the data quality after collecting.

Model requires daily meteorological data that could either be read from a measured data set or be generated by a weather generator model. In this study, daily precipitation, minimum and maximum air temperature used. These data were obtained from Ethiopian National Meteorological Agency (NMA) for stations located within and around the watershed.

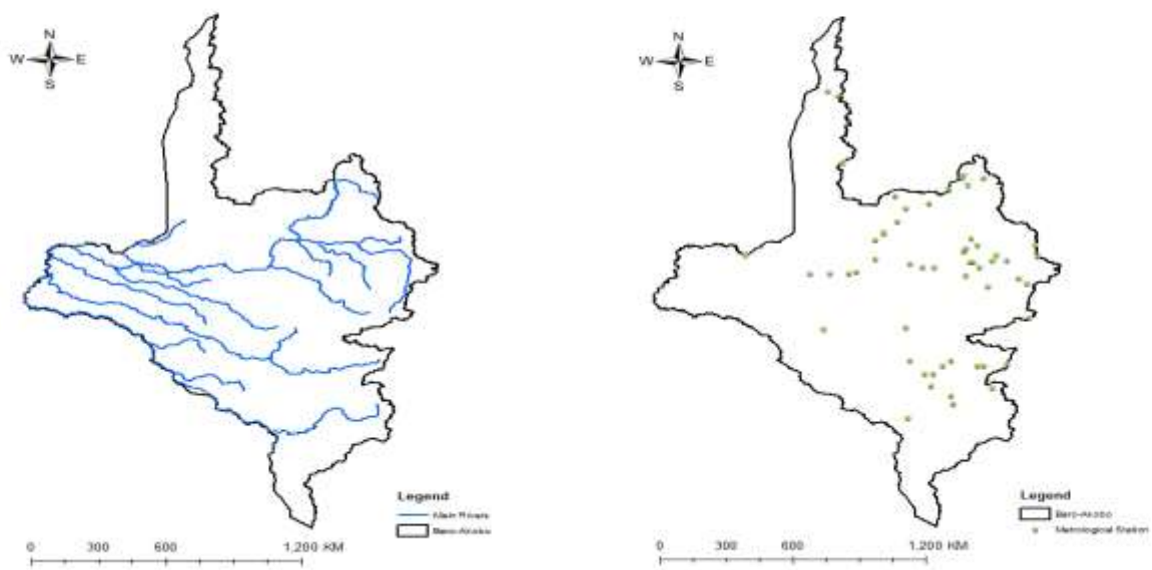

Figure 3:Main Rivers and Metrological Station

Stream flow data

The time series stream flow data was obtained used for the low flow analysis was obtained from the Ministry of Water Irrigation and Electricity. Among the functional stream flow stations, the data for low flow analysis from hydrological stations based on the long record of data, less missing and functionality. The stations consist of daily data were used to produce different duration of low flows and also used for return period.

Figure 4: Hydrological Stations

3.2.3 Remote sensing

Remote sensing data used in Landsat satellite imagery to measure the water surface width, longitudinal slope and water depth in river channels (Anette E., Hannes S. and Jens G., 2019). To estimate daily streamflow at an ungauged river will be done by the use of remote sensing data and Manning's equation (Manning R., 1889) in the following steps. The first three steps are processing of remote sensing data. After the data are prossed will be used in Manning's equation to give a standard stage–discharge relationship (rating curve) from which the estimated daily discharge is obtained.

$$Q = \frac{1}{n} R^{2/3} S^{1/2} A \qquad 1$$

In Step 1, channel width (W) measurements river geometry approach or at a specific location. The channel width measurements were recovered manually selected dates from cloud free images to cover a range of water levels, including typical low water levels and high-water levels. Landsat measurement dates are important later on in the method as they are the dates for which the average water depth (Y), the discharge (Q) and the stage–discharge relationship are estimated. Step 2 Calculation of a daily time series of estimated stage data water depth (d) using satellite altimetry data. Step 3, the longitudinal channel slopes (S) were estimated by selecting the satellite altimetry crossing (at the location where the discharge is to be estimated) and the next downstream crossing site. Step 4 involves estimating the unknown channel geometry below the minimum Landsat measurement to allow the calculation of the water depth (Y).

Step 4 Calculation of the unknown bathymetric depth (H_{min}), i.e. the stage level at which there is zero flow, and hence the calculation of average water depth (Y) on each of the Landsat measurement dates. Step 5, the discharge (Q) is using GR4J model estimated daily time series of discharge (Q). Step 6 is then calculated using the estimated daily time series of stage data (Step 2) and the estimated stage–discharge relationship in (Step 5). Note the estimated stage–discharge relationship is not strictly necessary. The daily time series of stage data could be used along with Manning Equation to give the daily time series of estimated discharge. However, using the stage–discharge relationship allows for extrapolation beyond the highest Landsat width, and fitting the curve reduces the errors associated with the daily time series of stage values. Roughness coefficient will be obtained from river characteristics.

3.2.4 GR4J (Ge´nie Rular 4` Parame`ter Journalier)

The GR4J model is a four-parameter daily lumped rainfall-runoff model, which belongs to the SMA (soil moisture accounting) models. The two inputs variables of the model are the precipitation depth P [mm] and the potential evapotranspiration PE [mm]; as first step the net rainfall Pn and the net evapotranspiration En are calculated. The four parameters are:The production store (X_1) is storage at the surface of the soil that holds rainfall. Evapotranspiration and percolation occur in this store according to soil moisture. The storage capacity depends on the types of soil in the river basin. Low porosity in the soil can increase the size of the production store. The groundwater exchange coefficient (X2) is a function of groundwater exchange, which influences the routing store. When X_2 has a negative value, water infiltrates to the aquifer; when it has a positive value, water exits the aquifer and adds to the routing storage. Routing storage (X_3) is the amount of water that can be stored in soil porosity. The value of X_3 depends upon the type and the humidity of the soil. The time peak (X_4) is the time when the ordinate peak of the flood hydrograph unit UH1 (day) is created during GR4J modelling.

The model is built up with a soil moisture (or production) store with an evolution described by the power function of its storage S; this first reservoir is followed by a transfer function that, in the daily version of the model, splits the water in two fluxes equivalent to 90% and 10% of the total. The first one is led to the unit hydrograph UH1 and then to a non-linear routing store, whereas the second skips the store, being routed directly by UH2. UH1 and UH2 are used to simulate the time lag between the precipitation and the streamflow peak (M. Andrea Ficchí., 2017). Moreover, a groundwater exchange function is applied to the two fluxes (Q9 and Q1) and is described by a power law of the routing store level R: finally, the two-resulting streamflow's Qr and Qd are summed up in order to obtain the final Q (Perrin, C. et al, 2015).

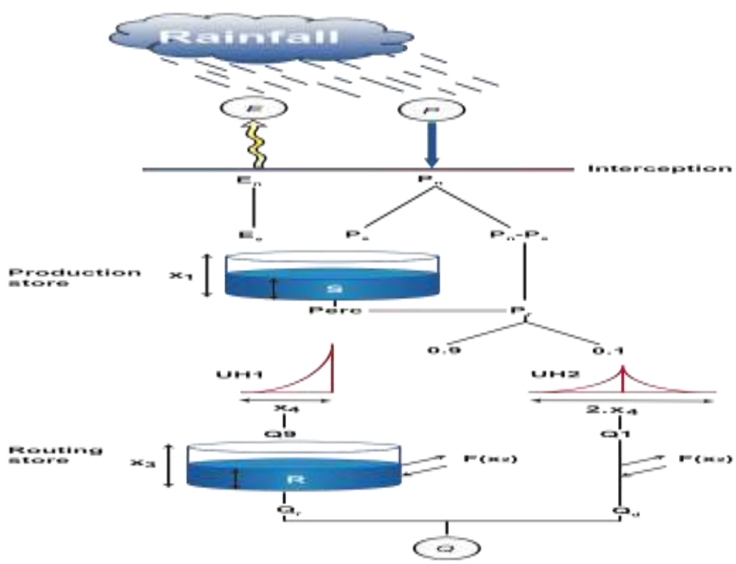

Figure 5: diagram of the GR4J model

3.2.5 Singular Spectrum Analysis

Singular Spectrum Analysis (SSA) technique consists of two stages known as decomposition and reconstruct and four steps known as embedding, singular value decomposing, grouping and diagonal averaging. SSA decomposes a time series and thereby enables differentiation between trend, harmonic and noise components, and then reconstructs a less noisy time series using the estimated trend and harmonic components, and this newly reconstructed series is then used to compute forecasts (Pietro B. et al., 2014). As a nonparametric method, SSA can be used without making any assumptions pertaining to stationarity and normality of the data. (Hassani, H., et al., 2013a). The SSA technique has both univariate and multivariate capabilities and applications of finding solutions to real world problems. The performance of the SSA technique depends upon the selection of its two parameters known as the matrix and number of eigenvalues.

The procedure performed during singular spectrum analysis can be described more formally in four main steps. Step one is the embedding of the time series in a high-dimensional lagged trajectory matrix. step two involves the decomposition of the trajectory matrix into the sum of a number of bi-orthogonal matrices of rank one. These two steps are considered together to be the decomposition stage. Next comes the reconstruction stage. Step three involves the summing of the various matrices that were formed in step two into different groups, depending on the nature of the matrices. Finally, in step four, the time series representing the various groups can be reconstructed from the resulting matrices (Nina G. and Anton K., 2013).

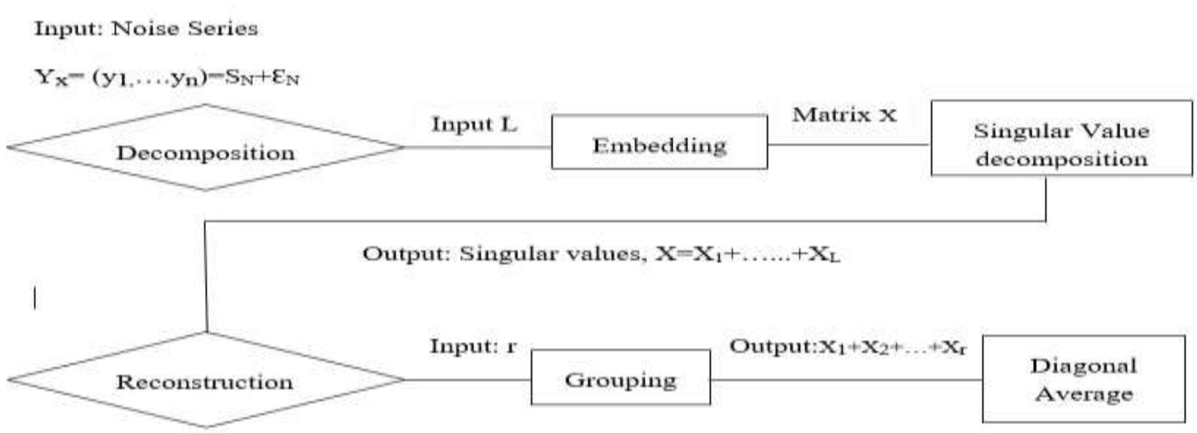

Figure 6: The Basic Process

3.2.6 Classification of low flow regimes for ungagged Rivers

Flow regime identification is an important application of regionalization and commonly the characteristics of streamflow quantity, timing and variability. Cluster analysis is the process of grouping similar catchments according to river location or river type (Ephemeral, Intermittent and Perennial Rivers). Fuzzy C-Means (FCM) is one of the most widely used fuzzy clustering algorithms applied to classify the flow regimes and identify the classification of groups. The fuzzy c-means clustering algorithm is a soft-clustering algorithm in which each data point assigned to a cluster of some degree that is specified by a fuzzy membership. Since the aim of our dataset clustering is a classification and cachacteristic of low flow regimes.

$$J = \sum_{i=1}^{1} \sum_{k=1}^{1} (\mu_{ik})^m \|X_k - C_i\|^2 \qquad 2$$

$$\mu_{ik} = \left[\sum_{i=0}^{n} \left(\frac{\|X_k - C_i\|^{2/(m-1)}}{\|X_k - C_j\|^2} \right) \right]^{-1} \qquad 3$$

$$C_i = \frac{\sum_{k=1}^{k} (\mu_{ik})^m x_k}{\sum_{k=1}^{k} (\mu_{ik})^m} \qquad 4$$

Here μ_{ik} is the membership coefficient of the k^{th} data object to the i^{th} cluster, m is the fuzziness coefficient greater than 1, X_k is the kth data object, C_i is the i^{th} cluster center, C is the number of clusters, and K is the number of data objects. The symbol || || denotes any vector norm that represents the distance between the data object and the cluster center.

3.2.7 Calibration

Model calibration involves modification of input parameters and comparison of predicted output with observed values until a defined objective function is achieved. Parameters identified in sensitivity analysis that influence the simulation result significantly were used to the model calibrate. Model calibration is a means of adjusting or fine-tuning model parameters to match with the observed data as much as possible, with limited range of deviation accepted. Sometimes it is necessary to change parameters in the calibration process other than those identified during sensitivity analysis because of the type of miss match of the observed variables and predicted variables (White K. L. and Chaubey I., 2005).

3.2.8 Validation

Validation is comparison of model results with an independent data set without further adjustment. The three statistical model performance measures used in calibration procedure were also used in validating stream flow. Similarly, model validation is testing of calibrated model results with independent data set without any further adjustment at different spatial and temporal scales.

3.2.9 Model Performance Evaluation

To find the values for the model parameters that minimize or maximize the specified calibration criterion and to formulate the regional model, level of goodness of fit is determined by the objective functions. In this study three types of objective functions are considered these are Coefficient of Determination (R^2), Nash-Sutcliff Efficiency (N_{SE}) and the percent difference (D). Coefficient of determination (R^2) describes the proportion of the total variance in the observed data that can be explained by the model. The closer the value of R^2 to 1, the higher is the agreement between the simulated and the observed flow and is calculated as follow:

$$\boldsymbol{R} = \frac{[\sum_{i=1}^{n}(\boldsymbol{Oi}-\boldsymbol{Oav})(\boldsymbol{Si}-\boldsymbol{Sav})]^{\wedge}2}{\sum_{i=1}^{n}(\boldsymbol{Oi}-\boldsymbol{Oav})^2 * (\boldsymbol{Si}-\boldsymbol{Sav})^{\wedge}2} \qquad 5$$

Where: Oi is measured value, Oav is average measured value, Si is simulated value, Sav is average simulated value, the same holds true for Equations (6) and (7). Nash and Sutcliffe simulation efficiency (E_{NS}) indicates the degree of fitness of observed and simulated data and given by the following formula.

$$\text{NSE} = 1 - \left[\frac{\sum_{i=1}^{n}(\text{Oi}-\text{Si})^2}{\sum_{i=1}^{n}(\text{Oi}-\text{Si})^2}\right] \qquad 6$$

The value of E_{NS} ranges from 1 (best) to negative infinity. If the measured value is the same as all predictions, E_{NS} is 1. If the E_{NS} is between 0 and 1, it indicates deviations between measured and predicted values. If E_{NS} is negative, predictions are very poor, and the average value of output is a better estimate than the model prediction (Nash and Sutcliffe, 1970). The percent difference (D) measures the average difference between the simulated and measured values for a given quantity over a specified period were calculated as follows:

$$D = 100 * \frac{\lfloor \sum_{i=1}^{n} o_i - \sum_{i=1}^{n} S_i \rfloor}{\sum_{i=1}^{n} o_i} \qquad 7$$

A negative value of D indicates model over estimation and a positive value indicate model under estimation.

4. EXPECTED OUTPUTS

In Baro Akobo river basin, most of subbasins/catchments are ungauged as a result there is less water development projects implemented so far in the basin as compared to other basins. Watershed rehabilitation and irrigation interventions are required to be implemented since soil erosion and sedimentation are the main challenges in majority of the watersheds. Therefore, estimating the stream flow in the ungauged watershed is essential for the purpose of designing and implementing infrastructure across the basin and to understand the different components of the hydrologic components. The estimated daily flow is applicable for low flow analysis only in ungauged or poorly gauged catchment areas needed for many decisionmakers in irrigation and water resources management. The planning, design, and management of effective water resource require a good estimate of streamflow. The derived flow estimation techniques and the application of the basin developing an ensemble approach for long-term low-flow forecasts on catchments distributed over the baro akobo basin. Low flow characteristics are usually estimated from flow gauge stations. However hydrological data are not always available at the site of interest: rain fall-runoff models, regionalization, remote sensing and singular spectrum analysis are commonly used for the estimation of flow characteristics at sites where little or no data exists.

5. Timeline and Budget Statement

5.1. 5.1 Timeline

Table 2: Time line

N0	Phases and Activities	Time Frame Work			
		1/09/2021-31/3/2022	1/4/2022-	1/9/2022-	1/2/2023-30/09/2023
1	**Data collection and Preparation**				
1.1	Field visit and discussion with concerned bodies				
1.2	Data collection; Metrological Data, Hydrological Data, and Remote sensing Data				
2	**Data analysis**				
2.1	Data preparing for models				
2.2	Data Quality Check,consistency test, homogeneity test, treating missing values Fill in missing data				
3	**Model**				
3.1	Rainfall-Runoff models; Model Calibration and validation and Parameter Transfer				
3.2	Remote sensing; Create River profile and Estimate surface water width				
4	**Final preparation and submission**				
4.1	Progress Report, Presentation,				
4.2	Three Journal preparation and publication				
4.3	Final work submission				

5.2. Budget

Table 3: Budget Statement

No	Items	Unit	Unit cost	Total Amount
1.	**PER DIEM: RESEARCHERS, ASSISTANT AND DAILY LABOR**			
1.1.	Per diem for preliminary survey (site investigation)	4 Researchers	4 researchers * 300Birr/day * 90days	108,000.00
	Per diem during all river's observation	4 Researchers	4 researchers * 300Birr/day * 120days	108,000.00
	Per diem during meteorological and other secondary data collection	4 Researchers	4 researchers * 300Birr/day * 120days	108,000.00
	Per diem for field assistants	4 Assistants	4 Assistants * 250 Birr/day * 80 days	80,000.00

	Labour cost for sample collection for geology/ hydrogeology study	4 daily labours	4 daily labours * 150 Birr/day * 120days	54,000.00
	Labour cost for sample collection for river cross section measurement	4 daily labours	4 daily labours * 150 Birr/day * 120days	54,000.00
	Sub Total			**512,000.00**
2. SERVICES: TRANSPORTATION AND LABORATORY				
2.1.	Car rent (land cruiser-vehicles-contract base 30 days)	1 car	1*30days*3000 Birr/day	90,000.00
2.2.	Public Transportation (researchers to town centres, within the field, taxi & sample transportation)	Lump sum	Lump sum	35,000.00
2.3.	Water samples analysis	4 seasons	3*110 sample point*200 Birr/sample	66,000.00
	Sub Total			**191,000.00**
3. Stationary cost				
3.1.	Personal Computer (HP Core i7)	1 Pcs	1*42000Birr	42000
	Flash disk	2 Pcs	2*600 Birr	1200
	External hard disk	2 tera	1*5000 Birr	5000
	Pens	2 Pack	50*2*8 Birr	800
	Note book	50 Note Book	50*15 Birr	750
	Paper	4 Pack	4*600 Birr	2400
	Printing	2000 Pages	2000*2 Birr	4000
	Color printing	600 Pages	800*5 Birr	4000
	Binding	5	5*350	1750
	Photocopying	2500 Pages	2500*1	2500
	Sub Total			**64,400.00**
4. EQUIPMENT & MATERIALS				
4.1.	**Rental equipment & materials**			
	Deep meter	1	1*120days*500 Birr/day	60,000.00
	Current meter	1	1*120days*500 Birr/day	60,000.00
	Sub Total			**120,000.00**
5	**Different services**			
5.1.	Data (Hydrology and Metrology Data Purchasing)	Lump sum	Lump sum	15000
	Software training	4 months	4*5000 Birr/Month	20000
	Journal Publication	4 Journals	4*15000Birr/Journal	60,000.00
	Sub Total			**80,000.00**
Total amount				**967,400.00**
Contingency 5%				**48,370.00**
Grand Total				**1,015,770.00**

6. REFERENCE

A. Loukas and L. Vasiliades. (2014). Streamflow simulation methods for ungauged and poorly gauged watersheds. *Natural and hazards Earth System Sciences*. doi:10.5194/nhess-141641-2014

A.R. Keshtkar. (2015). Low flow frequency analysis by L-moments method Case study: Iranian Central Plateau River Basin. *International Desert Research Center, University of Tehran, Tehran.*

Anette E., Hannes S. and Jens G. (2019). Flow velocity and discharge measurement in rivers using terrestrial and UAV imagery. *Hydrology and Earth since.* doi:.org/10.5194/hess2019-289

Arthur W. S et al. (2018). Estimation of River Discharge Solely from Remote-Sensing Derived Data: An Initial Study Over the Yangtze River. *Remote Sensing*.

Crochemore L. (2016). Seasonal streamflow forecasting for reservoir management. *PhD thesis, Irstea.*

Gregor L. (2014). Regional assessment of low flow processes and prediction methods across European regimes. *Hydrology in a Changing World. Montpellier, France.*

Hakan T. and Martijn J.B. (2018). Simulation and forecasting of streamflow using machine learning models coupled with base flow separation. *Journal of Hydrology*. doi: https://doi.org/10.1016/j.jhydrol.2018.07.004

Hassani, H., et al. (2014). A preliminary investigation into the effect of outlier(s) on singular spectrum analysis. *Fluctuation and Noise Letters, 13(4):1450029.*

Hassani, H., et al. (2013a). Forecasting before, during, and after recession with singular spectrum analysis. *Journal of Applied Statistics, 40(10):* 2290–2302.

K.L. Jaeger et al. (2018). Probability of Streamflow Permanence Model (PROSPER): A spatially continuous model of annual streamflow permanence throughout the Pacific Northwest. *Journal of Hydrology*.

L. Coron et al. (2017). The suite of lumped GR hydrological models in an R package. *Environmental Modelling & Software*.

M. J. Kirkby et al. (2011). Classifying low flow hydrological regimes at a regional scale. *Hydrology and Earth System Sciences*. doi:10.5194/hess-15-3741.2011

Manning R. (1889). On the flow of water in open channels and pipes. *Transactions of the Institution of Civil Engineers of Ireland 20:* 161-267.

N.T.S. Wijesekera. (2018). Classification of Streamflow Observations for Water Management. *Department of Civil Engineering, University of Moratuwa, Sri Lanka.*

Nash J. E. and Sutcliff J. V. (1970). River flow forecasting through conceptual models, part I a discussion of principles. *Journal of. Hydrology (10),* 282–290.

Ni. L.W. and Khin M.W. (2014). Low Flow Frequency Analysis on Selected Rivers in Malaysia. *International journal of Earth Sciences and Engineering. Vol:7 (04).*

Nina G. and Anton K. (2013). Basic Singular Spectrum Analysis and forecasting with R. *Computational Statistics and Data Analysis*.

Perrin, C. et al. (2015). Improved rainfall-runoff modelling tools for low-flow forecasting: application to French catchments.

Pietro B. et al. (2014). Singular Spectrum decomposition: A new method for time series decomposition. *World Scientific Publishing Company: Advances in Adaptive Data Analysis Vol. 6, No. 4.* doi:10.1142/S1793536914500113

Ruigang Z. et al. (2012). Baseflow separation in a small watershed in New Brunswick, Canada, using a recursive digital filter calibrated with the conductivity mass balance method. *Hydrological Processes, (27),* 2659–2665. doi:10.1002/hyp.9417

S. G. Gebrehiwot et al. (2011). Hydrological characterization of watersheds in the Blue Nile Basin, Ethiopia. *Hydro. Earth System Sciences.15,11-20,* doi:10.5194/hess-15-11-2011

Sang Ug K. and Kil S. (2010). Regional Low Flow Frequency Analysis Using Bayesian Regression and Prediction at Ungauged Catchment in Korea. *KSCE Journal of Civil Engineering 14(1),* 87-98. doi:10.1007/s12205-010-0087-7

Silva, E. S. and Hassani, H. (2015). On the use of singular spectrum analysis for forecasting U.S. trade before, during and after the 2008 recession. *International Economics, 141*, 34–49.

Singh, V.P. and Woolhiser D.A. (2002). Mathematical modelling of watershed hydrology. *Hydrology. Eng., 7(4),* 270-292.

Sustainet E.A. (2010). Technical Manual for farmers and Field Extension Service Providers: Soil and Water Conservation. *Sustainable Agriculture Information Initiative, Nairobi.*

Velpuri N. M. et al. (2012). A multi-source satellite data approach for modelling Lake Turkana water level: calibration and validation using satellite altimetry data. *Hydrology and Earth System Sciences.*

WMO. (2009a). World Meteorological Organization. Manual of low-flow estimation and prediction. *Operational hydrology report number 50, World Meteorological Organization (WMO) Report Number 1029. Geneva: WMO*.