AFRICA CENTER OF EXCELLENCE FOR WATER MANAGEMENT

ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES

QUANTIFYING UNCERTAINTIES OF REMOTE SENSING-BASED RAINFALL AND EVAPOTRANSPIRATION PRODUCTS FOR GROUNDWATER RECHARGE ESTIMATION IN THE TIKUR WUHA WATERSHED, RIFT VALLEY LAKES BASIN, ETHIOPIA.

By

TSEGAMLAK DIRIBA BEYENE

A PhD dissertation submitted to the Africa Center of Excellence for Water Management, the School of Graduate Studies of Addis Ababa University in partial fulfilment of the requirements for The Degree of Doctor of Philosophy in Water Management (Hydrology and Water Resource Management)

August 2024

AFRICA CENTER OF EXCELLENCE FOR WATER MANAGEMENT

ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES

QUANTIFYING UNCERTAINTIES OF REMOTE SENSING-BASED RAINFALL AND EVAPOTRANSPIRATION PRODUCTS FOR GROUNDWATER RECHARGE ESTIMATION IN THE TIKUR WUHA WATERSHED, RIFT VALLEY LAKES BASIN, ETHIOPIA

By

TSEGAMLAK DIRIBA BEYENE

A PhD dissertation submitted to the Africa Center of Excellence for Water Management, the School of Graduate Studies of Addis Ababa University in partial fulfilment of the requirements for The Degree of Doctor of Philosophy in Water Management (Hydrology and Water Resource Management)

DECLARATION

I, Tsegamlak Diriba Beyene (ACEWM/GSR/2796/12), hereby declaring that this research dissertation titled *"Quantifying Uncertainties of Remote Sensing-Based Rainfall and Evapotranspiration Products for Groundwater Recharge Estimation in the Tikur Wuha Watershed, Rift Valley Lakes Basin, Ethiopia."* has been developed by me and has not been submitted to any other institution for the award of any academic qualification. The content of the dissertation has not been plagiarized, and where the works of other researchers have been used, they have been appropriately cited.

Tsegamlak Diriba Beyene	_____	_____
Candidate's Name	**Signature**	**Date**

AFRICA CENTER OF EXCELLENCE FOR WATER MANAGEMENT

ADDIS ABABA UNIVERSITY

Quantifying Uncertainties of Remote Sensing-Based Rainfall and Evapotranspiration Products for Groundwater Recharge Estimation in the Tikur Wuha Watershed, Rift Valley Lakes Basin, Ethiopia

By

TSEGAMLAK DIRIBA BEYENE

A PhD DISSERTATION SUBMITTED
TO
AFRICA CENTER OF EXCELLENCE FOR WATER MANAGEMENT
ADDIS ABABA UNIVERSITY

APPROVED BY BOARD OF EXAMINERS

This is to certify that we have read this PhD research proposal and that, in our opinion, it is fully adequate, in scope and quality, as a PhD dissertation for The Degree of Doctor of Philosophy in Water Management (Hydrology and Water Resources Management).

Advisor
Name Dr. Fasikaw Atanaw Signature_____ Date_06 – 08 – 2024

Co-Advisor
Name Dr. Dessie Nedaw Signature_____ Date_06 – 08 – 2024

Co-Advisor
Name Dr. Sirak Tekleab Signature_____ Date_06 – 08 – 2024

Examiner
Name Prof Assefa Melesse Signature_____ Date_06 – 08 – 2024

Examiner
Name Prof Seifu Tilahun Signature_____ Date_06 – 08 – 2024

Chairperson
Name Signature_____ Date_06 – 08 – 2024

ACKNOWLEDGMENT

“Everything is possible with the will of God!” Thank you for giving me the strength to complete my study.

Above all, my gratitude goes to my supervisor Dr. Fasikaw Atanaw; your couching and advice encouraged me to work hard and complete my studies. Next, I would like to thank my coadvisors Dr. Dessie Nadaw and Dr. Sirak Tekleab, who provided me with great advice and support. Thank you for your tireless advice and critical review.

Additionally, I want to thank my dear friend Ashenafi Alemu Gemechu for supporting me in all that I do not have.

I would like to thank ACEWM for supporting my research and all the ACEWM staff for providing all the support I needed during my study.

Furthermore, my humble gratitude goes to my wife and daughter for standing beside me and tolerating me at my low points; this would not have been possible without your support. Finally, I want to thank my family’s friends who stood by me through this journey.

DEDICATION

Dedicated to my beloved wife Mulumebet Derese and my lovely daughter Anaf Tsegamlak.

You are my courage and strength; thank you for understanding my lowest behaviour during this study and for giving me courage by praising my success. This is for you!

Table of Contents

List of Tables

List of Figures

LIST OF ABBREVIATIONS

AP	Annual Precipitation
BATEA	Bayesian Total Error Analysis
BC	Bias Corrected
BFI	Base Flow Index
BL	Bilinear
BSF	Base Flow Separation Method
BWB	Basin Water Balance Method
CCD	Cold Cloud Duration
CHIRP/S	Climate Hazards Group Infrared Precipitation
CM	Chloride Mass Balance
CM	Conditional Merging
CMB	Chloride Mass Balance
Cv	Coefficient Of Variation
DEM	Digital Elevation Model
DM	Distribution Mapping
EMA	National Ethiopian Meteorological Agency
ET	Evapotranspiration
FA	False Alarm/Precipitation
FAR	False Alarm Ratio
FBI	Frequency Bias Index
G.C.	Gregorian Calendar
GDA	Geographic Differential Analysis
GDAL/OGR	Geospatial Data Abstraction Library
GLEAM	Global Land Evaporation Amsterdam Model
GLUE	Generalized Likelihood Uncertainty Estimator
GRA	Geographical Ratio Analysis
GUI	Graphical User Interface
GWR	Geographically Weighted Regression
HB	Hit Bias
IBUNE	Integrated Bayesian Uncertainty Estimator
IPS	Isotope Peak Shift
ITCZ	Inter-Tropical Convergence Zone
LSTM	Long Short Term Memory
MAP	Mean Annual Precipitation
MB	Miss Bias
MCD+N	Monte Carlo Dropout With An Input-Dependent Data Noise Term
MCMC	Marcov Chain Monte-Carlo
MER	Main Ethiopian Rift
MODFLOW	Modular Finite-Difference Flow Model

MODIS	Moderate Resolution Imaging Spectroradiometer
MRC	Master Recession Curve
MVE	Model Volumetric Efficiency
NA	Not Applicable/Available
NETCDF	Network Common Data Form
NM	Numerical Modelling
NN	Nearest Neighbor
NSE	Nash Suit Cliff Efficiency
NUMPY	Numerical Python
OK	Ordinary Kriging
PANDAS	Python Data Analysis
PEST	Parameter Estimation And Uncertainty Analysis
POD	Percent Of Detection
PVE	Percent Error In Volume
QM	Quantile Mapping
RMSE	Root Mean Squared Error
SA	South Africa
SBCM	Spatially Bias Corrected And Conditionally Merged
SCIPY	Scientific Python
SFPT	Secure File Transfer Protocol
Smroot	Root-Zone Soil Moisture
Smsurf	Surface Soil Moisture
SRE	Satellite Rainfall Estimates
SRP	Satellite Rainfall Product
STDEV	Standard Deviation
SWAP	Soil Water Atmosphere Plant
SWAT	Soil Water Assessment Tool
SWB	Soil Moisture Balance Method
TOPMODEL	Topography Based Hydrological Model
TPRDF	Two-Parameter Recursive Digital Filter
UNSAT	Unsaturated Soil Water And Heat Flow Model
UTM	Universal Trans-Mercator
WAVES	Water Atmosphere Vegetation Energy And Solutes
WTF	Water Table Fluctuation

ABSTRACT

Groundwater recharge estimation is important for better groundwater management and decision making. Despite advancements in methods for groundwater recharge estimation and the availability of different methods, the results acquired are still variable and uncertain. Due to its hydrogeological formation, slight recharge variability in the Hawassa basin causes significant groundwater level variability. Therefore, this study focused on investigating groundwater recharge estimation uncertainty caused by input data in the Tikur Wuha watershed. Additionally, this study aimed to investigate the effectiveness of rainfall data assimilation from ground observations and climate hazards group infrared precipitation (CHIRP/S) satellite rainfall estimates (SREs) to decrease the uncertainty of recharge estimations. Accordingly, using a conditional merging procedure, two versions of resampled and spatially bias-corrected CHIRP estimates were merged with ground measurements. Additionally, the physically based fully distributed hydrological model WetSpa was used to simulate 30,000 possible combinations of parameters (i.e. randomly generated through Monte Carlo simulation stratified by Latin hypercube sampling (LHS)) for the three model setups. The M1 model setup was developed based on the rainfall measurements obtained from rain gauge stations scattered in and around the Tikur-Wuha watershed in Ethiopia, and the M2 model setup was developed using bias-corrected SREs of CHIRP merged with relevant ground station records. In contrast, the M3 model setup was executed using version 3.8a GLEAM evapotranspiration. One hundred best-performing parameter combinations were selected for each model setup to generate spatial recharge statistics and assess the resulting uncertainty in the recharge estimates. The results of the applied performance measures (i.e., seven) on the corrected and merged CHIRP SREs show that the percentage of detection (POD) and percent volume error (PVE) improved. Moreover, over the sparsely populated western part of the Lake Hawassa basin, the bias-corrected and conditionally merged CHIRP SREs outperformed the estimates obtained by CHIRPS. However, the devised multistage bias correction was limited in considering dry-day events during bias correction, which affected the bias correction performance of the CHIRPS product. On the other hand, the results of the uncertainty assessment revealed that enhanced spatial recharge estimates can be produced through improved CHIRP-based SREs. The replacement of ET estimates using ground meteorological records with the GLEAM dataset reduced the Cv value by 54% compared to M2. However, uncontrolled irrigation water uses and total recharge from

irrigation fields scattered across the Tikur-Wuha watershed were not considered in the study, which is a limitation of the study. Finally, future research should concentrate on methods of fusing to understand the benefits of various approaches and produce more precise rainfall records. Additionally, future studies should consider the contribution of irrigation water to the total recharge of the watershed to analyze recharge uncertainty.

CHAPTER ONE

1.0. Introduction

1.1. Background of the study

Groundwater recharge estimation is important for better groundwater management and decision making. Recharge is considered the main controlling factor of subsurface processes in groundwater management and related decisions (Mohan et al., 2018). Despite advancements in the methods of estimation and the availability of different methods, the results acquired are still variable and uncertain(Cheng et al., 2018). In addition to the recharge process, other subsurface phenomena, such as interbasin groundwater net flux, abstraction, evapotranspiration, effluent and/or influent interaction with surface storage, are considered complex in analysing the water balance of subsurface storage (Berehanu et al., 2017). Therefore, input data uncertainty, in addition to model parameter uncertainty and model structural uncertainty, cannot be completely avoided; rather, it can be minimized.

The reliability of estimation can be increased by implementing multiple methods of estimation (Scanlon et al., 2002). However, such methods provide a synoptic view of the uncertainty rather than providing a source of uncertainty (i.e., uncertainty of the input or parameter or model structure) related to the implemented methods. Determining the uncertainty of a specific method has been emphasized by different researchers. Delottier et al. (2018) assessed the uncertainty of the water table fluctuation (WTF) method by analysing the uncertainty of specific yields from aquifer tests, which were subsequently propagated for recharge estimation. The results indicated that the majority of the uncertainty in the WTF method emanates from the uncertainty of specific yield estimations (Delottier et al., 2018). The conventional chloride mass balance method has also been investigated for model structure uncertainty from the enhancement of canopy chloride deposition (Deng et al., 2013), chloride mass balance (CMB) model structural uncertainty for regions with dominant point recharge (Somaratne and Smettem, 2014), and spatial input variable uncertainty due to the upscaling of point estimation (Crosbie et al., 2019, Ordens et al., 2012, Alcalá and Custodio, 2014), whereas the uncertainty due to CMB parameter uncertainty was studied by (Crosbie et al., 2018).

The basin water balance method is reported to yield rough estimates (Berehanu et al., 2017, Adomako et al., 2010), where it is usually underestimated. This could partly be attributed to the fact that conventionally, the method will be simplified by assuming that components such as withdrawal, groundwater inflow/outflow, and sometimes evapotranspiration from deep groundwater storages are negligible. However, such a rouge assumption could lead to greater uncertainty in the estimation of recharge. In

previous studies such as (Alcalá and Custodio, 2014, Crosbie et al., 2019, Ordens et al., 2012, Xie et al., 2017, Xie et al., 2018), the input data error propagation method has been implemented to assess the uncertainty in recharge estimation. Berehanu et al. (2017) used a water balance technique to determine the groundwater inflow to the upper Awash Basin from adjacent Blue Nile subcatchments. The study concluded that the recharge contribution from adjacent catchments could have a greater influence on the accumulated recharge in the focus basin.

Basin water balance analysis has been widely used to investigate the different dynamics of the hydrological cycle and component interactions (Tekleab et al., 2011). An investigation of the uncertainty related to the estimation or data acquisition of these components can provide new insight into the contribution to the output (in this case, recharge). Xie et al. (2018) argued that while using water balance analysis for the estimation of potential groundwater recharge, the related uncertainty could be reduced significantly if one can produce more accurate field-based evapotranspiration time series data. These findings indicate that the use of high-resolution (i.e., spatial and temporal) meteorological data could improve the accuracy of hydrological computations (Müller-Thomy et al., 2018). On this note, for a data-scarce region with a paucity of meteorological stations, the spatial uncertainty will increase. However, compared to the increase in evapotranspiration data from satellite-based and reanalysis estimates, the impact of using alternative evapotranspiration data has not been investigated.

The uncertainty of meteorological data, in addition to the unevenness of the station distribution, can arise from equipment errors and data collector errors. The severity of the problem will increase in developing countries (Dile et al., 2018). Even though data acquired from remote sensing have their own errors (Gebremedhin et al., 2021), it is expected that some of the ground observation uncertainty sources (such as spatial coverage) are compensated, and the reverse is true. Data assimilation techniques are the best tools for obtaining the best futures of both methods (Gebremedhin et al., 2021). Such approaches not only provide better outputs but can also be used to analyse the related uncertainties and provide results with some probability.

Merged in situ rainfall observations with satellite rainfall products were evaluated by Gebremedhin et al. (2021) in the upper Tekeze Basin, Ethiopia. In this research, the geographical weighted regression method was used to merge two satellite rainfall products with available ground observations. This result indicates that better performance was obtained from the merged MPEG product than from the merged CHIRPS product. Another study by Westerhoff et al. (2018) introduced a large-scale recharge estimation model for New Zealand by using high-resolution satellite data and ground observation maps. The study concluded that ground observation rainfall data and geological maps lead to the highest uncertainty. As far as the literature review conducted for this study indicates, the common gap of such studies is that very little has

been done to investigate merged products for small catchments with data scarcity. Similar data scarcity and uncertainty exist in the focus basin (Bitew and Gebremichael, 2011). Nevertheless, the use of a data assimilation technique to assess and minimize the uncertainty caused by meteorological data for recharge estimation using numerical water balance modelling has not been well investigated.

Very few studies have conducted vague recharge uncertainty analyses in Ethiopia; however, these studies considered uncertainty due to the use of input data and the creation of recharge maps. Therefore, the main objective of this study is to quantify the uncertainty in recharge estimation due to input meteorological data. Additionally, this study aimed to investigate the effectiveness of assimilating ground observation and remote sensing data to decrease the uncertainty of recharge estimations.

1.2. Problem Statement

Groundwater supplies drinking water for 25% to 40% of the world's population (Vrba and van der Gun, 2004) and is the largest food supply produced by irrigation, accounting for approximately 38% of irrigated areas worldwide (Siebert et al., 2010). The cultivated area covers 62% of the lake Hawassa basin (Degife et al., 2019), of which a considerable percentage is irrigated. It is expected to provide a sustainable groundwater supply to major cities along the Addis Ababa-Nairobi-Mombasa road corridor. In addition to anthropogenic stress, the Lake Hawassa catchment is considered to have considerable groundwater outflow (Ayenew, 2001, Ayenew and Tilahun, 2008, Tessema, 2004). Therefore, sustainable management of this resource is crucial for maintaining a reliable groundwater supply.

Recharge has been given more emphasis because it is considered to be the main controlling factor in sustainable groundwater resource management (Ebrahim et al., 2020). Although there is a reservation on its applicability (Gorelick and Zheng, 2015), the general rule of conservation for sustainable use is to limit the extraction rate by the natural recharge rate. Additionally, studying the uncertainty of recharge is beneficial for prioritizing land use management tailored to increasing groundwater recharge. Another urging matter requiring investigation of recharge uncertainty is the growing risk of groundwater contamination (Abdurahman and Zewdie, 2018, Feng et al., 2015), which is due to urban population growth (Degife et al., 2019) and industrial zone expansion in the basin (Yeneneh, 2014). Since the basin has no surface outflow, contaminates generated in the basin either join the lake (eventually join the groundwater) or directly leach to the groundwater (Yeneneh, 2014). Additionally, slight recharge variability in the Hawassa Lake basin causes significant groundwater level variability (Ayenew and Tilahun, 2008). From stream networks in the Lake Hawassa basin, the Tikur Wuha River needs further investigation since it drains 50% of the catchment area (Degife et al., 2019) and is the most populated part of the watershed, generating more contaminants (Abiye, 2008). Therefore, it is essential to understand the ranges of uncertainty in recharge estimation methods.

Previous recharge estimates for the Lake Hawassa watershed (Table 1) have shown significant discrepancies. The estimated minimum recharge ranged from 50 mm/year to 190 mm/year (Seifu, 2013), while others have suggested that the actual recharge ranges from 17.885-80.3 mm/year (Ayenew and Tilahun, 2008). Such variation in the estimation indicates uncertainty in the estimation. Despite the efforts made to quantify recharge in the basin (Ayenew and Tilahun, 2008, Lemlem, 2008), significant uncertainty could exist due to the scarcity of input data, model parameter variability, or model structure uncertainty. Modelling without uncertainty analysis is still possible, but it is better presented at the uncertainty level (Xie et al., 2018). Therefore, this study focused on investigating the uncertainty caused by input meteorological data on groundwater recharge estimates for the Tikur Wuha watershed of the Lake Hawassa Basin, Ethiopia.

1.3. Significance of the Study

A major land use change in the Tikur Wuha River catchment is the increase in agricultural area (Abraham et al., 2022), which includes a large area of irrigation fields. Some of these privately owned irrigation fields receive water from groundwater (Ayenew and Tilahun, 2008). Hence, proper management of the extraction rate in relation to the actual recharge and the level of estimate is essential for sustainable management of groundwater. Usually, due to its complexity, the need for estimation is limited to the range of recharge, i.e., the lower and higher limits. The lower limit represents the actual recharge, and the potential recharge is the higher limit. In recharge estimation, it is never easy to use the term ‘accurate’ to indicate the level of estimation (Healy, 2010). Therefore, it is rather better explained with uncertainty analysis. Here, uncertainty is inferred from the parameters, model structure, model input data or model initialization and boundary conditions (Ajami et al., 2005, Moges et al., 2021). Nevertheless, the recharge estimations for a given area should be presented with an uncertainty level or reliability level of estimation.

Ensuring the reliability of the estimate could benefit decision-making in different ways. For instance, successful implementations of managed aquifer recharge projects rely on accurate spatial and temporal estimates of recharge. Such practices are very helpful in creating a groundwater resource that is resilient to climate change and variability (Ebrahim et al., 2020). As most contaminant sources are diffuse, reliable spatial estimates of recharge are useful for the management of head waters upstream of extraction wells(Kurylyk and MacQuarrie, 2013, Ebrahim et al., 2020). Nevertheless, most recharge estimates lack an uncertainty analysis. Xie et al. (2018) found that the main reason, in addition to the complexity of the analysis, is that recharge estimates are used relatively rather than in an absolute sense. Nevertheless, the importance of uncertainty analysis of the recharge estimation method is undeniable. Therefore, this study aims to investigate the contribution of input uncertainty to recharge estimates.

In conclusion, this study will help to investigate the uncertainty of groundwater recharge estimates due to input rainfall and evapotranspiration data. In an attempt to investigate alternative data sources in addition to the conventionally used records of ground meteorological stations, this study evaluated a new multistage bias correction approach for CHIRP/S satellite rainfall estimates (SREs). Additionally, the use of a satellite-based reanalysis evapotranspiration product known as GLEAM was evaluated for uncertainty reduction.

1.4. Research Objective

The main objective of this study is to quantify the uncertainty in recharge estimation due to input meteorological data in the Tikur Wuha River catchment in the Rift Valley Lakes Basin, Ethiopia.

The specific objectives of this study are as follows:

- To bias correct the CHIRP and CHIRPS satellite rainfall products coupled with the conditional merging technique
- To examine the uncertainty of groundwater recharge using merged satellite rainfall products with ground observations.
- To evaluate the contribution of uncertainty due to evapotranspiration to the estimation of groundwater recharge

1.5. Research Questions

- Can we improve the quality of rainfall maps by merging bias-corrected CHIRP SRE with ground observations?
- Is the performance of bias-corrected and conditionally merged CHIRP SRE superior to CHIRPS SRE?
- Does recharge uncertainty decrease when using merged rainfall products?
- Does the uncertainty of groundwater recharge estimates improve for different evapotranspiration data?

1.6. Scope of the Study

The designed study is limited to investigating the uncertainty of groundwater recharge estimation due to input rainfall and evapotranspiration data. CHIRPS and its prior version, the CHIRP satellite rainfall product, are the only satellite-based rainfall estimates investigated in this study, whereas version 3.8a of the GLEAM potential evapotranspiration (ET) data was the only satellite-based estimate of ET evaluated. Additionally, the study is limited in its ability to investigate the uncertainty by only taking volumetric performance measures, where the modelling iterations only consider the resulting RMSE and MVE. Hence, the resulting mean annual recharge estimates cannot be taken as the actual or near-actual estimates for the

Tikur-Wuha watershed. To include sufficient ground meteorological station records, multistage bias correction was applied to the Hawassa basin, whereas an uncertainty assessment was conducted for the Tikur-Wuha watershed. Furthermore, a calibration of the WetSpa model was conducted only for the Wosha sub-watershed, which is one of the sub-watersheds of the Tikur-Wuha watershed. Finally, the study was limited to including the impacts of irrigation inputs and other manmade abstractions.

CHAPTER TWO

2.0. Literature review

2.1. Review of uncertainty analysis

Over the years, different methods have been introduced to quantify recharge. However, it should be noted that a recharge definition is necessary whenever a particular method is referenced. Scanlon et al. (2002) categorized the methods based on the hydrological source/zone as the saturated zone, unsaturated zone, and surface water. The classification was further subdivided into physical methods, tracer methods, and numerical methods (Scanlon et al., 2002). Such classification may depend on the purpose of classification; for instance, the above classification was made with the aim of distinguishing methods based on spatial and temporal ranges with associated reliability (Scanlon et al., 2002). Additionally, recharge estimation from all zones can also be achieved using methods such as basin water balance, numerical modelling, and empirical rainfall-recharge relationships (Walker et al., 2019).

Inherent to their assumptions and structural nature, each method is limited both spatially and temporally, and they can only detect a range of recharge magnitudes (Scanlon et al., 2002). For further details, readers are referred to Healy (2010), Lorentz et al. (2003), and Scanlon et al. (2002). Numerous studies have been conducted on the uncertainty analysis of different recharge estimation methods. In this review, special focus has been given to the methods applied most frequently. The uncertainty in recharge estimates is sometimes due to a lack of understanding of the assumptions for a specific method or to not meeting the data requirements and errors created during data collection (Healy, 2010). However, such sources of uncertainty will not be covered in this review, as these types of errors cannot be used to deduce the method.

2.1.1. Tracer-based estimation methods

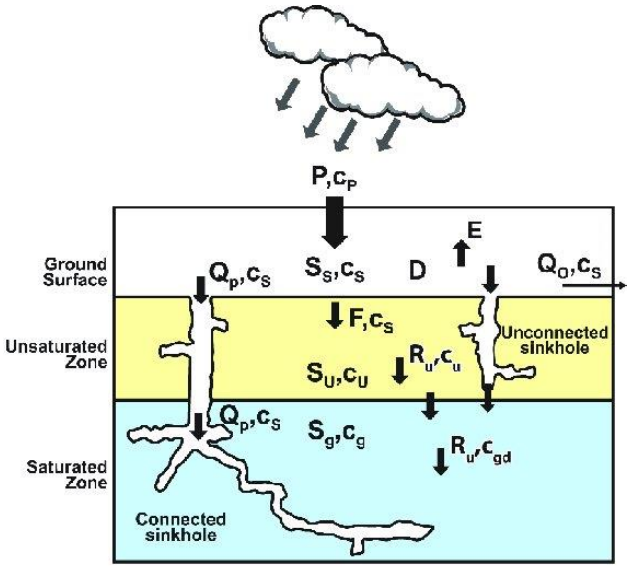
The general rule of tracer application for estimating recharge (i.e., mass balance approach) follows the fractionation of the concentrations of selected tracer elements from different sources. This means that the concentration of the tracer element will be measured in the rainfall and will be compared to the concentration in the groundwater or other points of measurement, such as springs. Tracer-based groundwater recharge estimation methods can be classified further into experimental (applied tracer), historical and environmental trace element methods (Healy, 2010). The experimental method usually depends on the judgment of the researcher, which implies that to investigate the uncertainty related to this method, one should collect all the assumptions, field setups, data collection processes, and other

factors(Meyer et al., 1997). These could vary from research to research, making the generalization of the method difficult. Therefore, this review discusses the uncertainty related to the application of environmental tracer elements. In the following section, widely applied methods such as CMB and stable isotope profiling are discussed in detail.

Model structure uncertainty: The chloride mass balance (CMB) method is widely applied (Scanlon et al., 2006) and tested for uncertainty in different parts of the world. The conventional method estimates diffuse recharge by assuming that the only source of chloride is the environment (Healy, 2010). Essentially, this means that point recharges and other sources of chloride are ideally negligible. The method's mathematical representation (Eqn. 1), as referenced by Marei et al. (2010), is:

$$\boldsymbol{R = P * \frac{Cl_p^-}{Cl_{gw}^-}} \text{ ---------- } \textit{Eqn. 1}$$

where R is the recharge; P is the precipitation; and Cl_p^- and Cl_{gw}^- are the chloride concentrations in the precipitation and ground water, respectively. It is indubitable that the above simple model will have uncertainty if other sources of chloride (Figure 1) are not included.



Note: Ss, surface storage; **Cs**, $\boldsymbol{Cl^-}$ concentration of surface storage; **P**, average annual precipitation; **Cp**, $\boldsymbol{Cl^-}$ concentration of rainfall; **D**, dry deposition; **E**, evapotranspiration; Qp, runoff to sinkholes; **Qo**, runoff out of catchment; **F**, infiltration; **Su**, unsaturated storage; **Cu**, $\boldsymbol{Cl^-}$ concentration in recharge water; **Ru**, diffuse recharge through the soil profile; **Sg**, groundwater storage; **Cg**, $\boldsymbol{Cl^-}$ concentration of groundwater; **Cgw**, $\boldsymbol{Cl^-}$ concentration of diffused groundwater recharge.

Figure 1 Schematic diagram for applying chloride mass balance to a control volume (Somaratne and Smettem, 2014)

However, one could imagine that the cost and time required to collect and analyse all sources are not compatible with the results of reducing (not eliminating) CMB model uncertainty. For regions where concentrated recharge cannot be neglected, such as in karst aquifers (Ahmet et al., 2020), updates have been made to include point recharge (Somaratne and Smettem, 2014). Deng et al. (2013) assessed the uncertainty of the traditional CMB model by considering the effect of the canopy on the surface deposition of chloride. The results indicate 28% and 89% chloride deposition enhancement for eucalyptus- and pine-dominant areas, respectively, compared to open fields. This clearly indicates the uncertainty in the model structure.

Some researchers argue that stable water isotopes are good indicators of recharge processes and sources but are less efficient when used for estimation (Scanlon et al., 2002, Walker et al., 2019). Nevertheless, the analysis has been applied for both purposes and has been reported to contain uncertainty. The uncertainty of this method when applied for source identification or process understanding is limited when there are multiple sources (Solder and Beisner, 2020). This means that different sources can have similar isotopic signatures. In such cases, it is recommended to use multiple types of isotopes and apply mixing models (Yu et al., 2018). Such types of models can sometimes be sources of uncertainty (Zhang et al., 2018b, Vázquez-Suñé et al., 2010, Solder and Beisner, 2020), but in general, the more isotope types that are used, the smaller the model uncertainty is.

Other forms of tracers include historical tracers, which are used to estimate the age of groundwater while inferring its recharge (McCallum et al., 2015). In the review by McCallum et al. (2015), the use of lumped parameter models to derive mean transit times and multiple tracers reduced model uncertainty. In contrast, Turnadge and Smerdon (2014) argued that numerical transport modelling allows direct and biased comparisons with observations, unlike other methods, which require apparent age bias correction.

Input data uncertainty: The CMB method provides point estimation for a given location, whereas recharge is a spatial phenomenon. As cited by Crosbie et al. (2018), pioneering studies assumed the estimate to be representative of an area, but later works such as Bresciani et al. (2014) argued that the estimate was an integration of an area upgradient of the sample location. This leads to spatial uncertainty in the process of upscaling point estimates. A study conducted in the Uley South Basin, South Australia, using water table fluctuation (WTF), CMB and isotope analysis indicated the uncertainty level by providing the range of estimates (i.e., high and low) and attributed the uncertainty to the input data (Ordens et al., 2012). Ordens et al. (2012) concluded that decreasing the uncertainty from other sources of chloride, the influence of vegetation, and proximity to the ocean could reduce such uncertainty.

Alcalá and Custodio (2014) suggested that using short-term chloride concentration records (1 to 5 years) will induce uncertainty in the estimation of long-term average recharge. In this study, the input variable

error propagation method was used to estimate the natural uncertainty and spatial uncertainty using the ordinary kriging upscaling technique. The study also concluded that better data coverage reduces the upscaling uncertainty. Davies and Crosbie (2018) used 291 measurements and a regression (chloride concentration and distance to the coast) in data-scarce areas. Then, ordinary kriging interpolation was applied to create a continental chloride concentration map for Australia. The uncertainty analysis was performed during the calibration by creating 1000 maps using the Monte Carlo sampling technique. The uncertainty analysis suggested that greater uncertainty was observed in regions where data were scarce, which supports Alcalá and Custodio (2014)the findings of .

Parameter uncertainty: The CMB method is susceptible to uncertainty generated from the unknown long-term deposition of chloride in the atmosphere, even though one could manage to meet all the assumptions and requirements of the method (Custodio, 2010). This could also be because of the chloride concentration in surface runoff, and the chloride concentration in groundwater is not known. Crosbie et al. (2018) improved scaling point estimation by using regression kriging and applied global regression using gridded rainfall with geology in data-scarce areas. The parameter uncertainty was analysed using 1000 stochastic replicates of the chloride deposition of rainfall, the chloride exported in runoff, the chloride concentration of the groundwater and the regression equations used to perform the upscaling. Additionally, a qualitative uncertainty analysis was performed to ensure that the assumptions of the method were consistent for the study basin.

On the other hand, one of the reasons why stable isotope performance is minimal in recharge estimation is linked to the susceptibility of ions to sorption, evaporation, dispersion and adsorption or other geochemical or biochemical processes (Healy, 2010, Cartwright et al., 2017). Adomako et al. (2010) applied three estimation methods, including the peak shift method, where the difference in the groundwater isotope peak is compared to the seasonal rainfall peak. The main assumption of this method is that the vertical permeability is much greater than the horizontal permeability, which neglects dispersive flow (Adomako et al., 2010). Soil depth profile samples from selected representative locations were analysed to acquire the hydraulic parameters of the depth profile and isotope profile of the soil water, where the results were used to profile additional requirements (hydraulic conductivity and isotope dispersivity) for transient flow and transport modelling. The uncertainty analysis was conducted by varying the hydraulic conductivity and other soil retention parameters, and the results exhibited >5% uncertainty.

2.1.2. Numerical modelling

Numerical modelling methods are applicable for estimating recharge in all hydrological zones (i.e., surface water, unsaturated zones and saturated zones) (Walker et al., 2019). Furthermore, they provide different scale recharge estimations, such as point scales from simulations of water flow through unsaturated zones (e.g., HYDRUS-1D (Simunek et al., 2005), UNSAT (Fayer, 2000), and SWAP (Kroes et al., 2017)) or large scales from catchment models (e.g., SWAT (Neitsch et al., 2002) and MIKE-SHE (Refshaard and Storm, 1995)).

Input data uncertainty: Despite their range of applicability, these methods need to be checked for reliability using field observations (such as lysimeter data, tracers, water content, and temperature/evaporation) (Scanlon et al., 2002). A common approach for addressing the uncertainty of these methods is to assess the uncertainty of the input data and deduce the uncertainty of the output. The spatial uncertainty of recharge estimation can be estimated by creating gridded input data or by creating similar regions of input data (Xie et al., 2018, Szilagyi et al., 2012).

Xie et al. (2017) argued that while using water balance analysis for the estimation of potential groundwater recharge, the related uncertainty could be reduced by 50% if one can produce more accurate field-based evapotranspiration time series data. In contrast, he reported that the use of field-measured moisture time series to reduce the uncertainty in recharge estimation is unsatisfactory. Xie et al. (2018) performed an uncertainty analysis of recharge estimates using the WAVES model (Zhang and Dawes, 1998) to compare modelled evapotranspiration to MODIS evapotranspiration data by applying the Monte Carlo method together with the Latin-Hypercube sampling technique to 13 predefined zones. Each zone was represented by the WAVES model for unsaturated zone analysis. This study highlighted that the MODIS ET can be used to reduce uncertainty. Another study by Westerhoff et al. (2018) introduced a large-scale recharge estimation model for New Zealand by using high-resolution satellite data and ground observation maps. The study achieved a 17% uncertainty level, and ground observation rainfall data and geological maps led to the highest uncertainty. The method implemented in the analysis of the uncertainty was propagation of variance and covariance of all input data (Westerhoff et al., 2018).

Model structure uncertainty: Many studies have been performed on the uncertainty of numerical models due to input data uncertainty as well as parameter uncertainties. Even though the uncertainty of the model structure could match the combined effect of the abovementioned uncertainties, little has been done to investigate model uncertainty (Højberg and Refsgaard, 2005). To some extent, this problem is related to data requirements and methods of analysis. The land cover type is one of the controlling factors of recharge. Ampe et al. (2012) used two land use classification approaches and implemented Monte Carlo simulation

to assess the uncertainty due to the classification method. The results illustrated that regional classification yielded lower uncertainty than pixel-based classification.

Ye et al. (2010) applied the model averaging method to assess model uncertainty using five recharge estimation methods. The recharge controlling factor geology was combined with these five models, resulting in 25 discrete models. The study concluded that the model uncertainty was greater than the parameter uncertainty. Model uncertainty induced during the calibration of parameters (parameterization process) was assessed by Zhu et al. (2020). In this study, hydrological simulation was conducted using three models, and the RORA model (Rutledge, 2007) was used for recharge estimation, whereas the uncertainty interval and coverage probability were used to assess the uncertainty. The results showed that parameterization can be a source of uncertainty when estimating diffuse recharge from simulated stream flow.

Manna et al. (2019) implemented water balance analysis using a spatially distributed numerical model (MIKE SHE (Refshaard and Storm, 1995)) to map the spatiotemporal distribution of recharge in an upland catchment in a fractured sandstone upland catchment near Los Angeles, California. In his investigation, he incorporated the analysis of stable water isotopes as a validation of his result. Qualitative uncertainty analysis was conducted by comparison with the analysis results from water isotopes. The research also concluded that using fine gridded (and fine temporal resolution) data will necessarily reduce the model uncertainty level.

2.1.3. Physical-based methods

Over the years, multiple physical-based methods have been developed to investigate groundwater recharge. Widely practiced methods involving the water table fluctuation method and base flow separation method are reviewed and discussed in the following section.

Parameter uncertainty: The main cause of uncertainty in the water table fluctuation (WTF) method is the uncertainty of the specific yield (Jeong and Park, 2017, Healy and Cook, 2002). There are many methods for determining the specific yield presented in the literature. Readers are referred to Healy (2010) for more details. One of these methods for investigating specific yield is soil core sample analysis. Even though this method is expensive, especially for heterogeneous geology, it is a reliable method with minimal uncertainty (Kim et al., 2010). Delottier et al. (2018) used aquifer tests with the water table fluctuation method to investigate the uncertainty due to the parameters of the method. The experimental setup in this study consisted of one pumping well at a distance of 6.8 m from this well. The uncertainties in the estimated specific yield and effective water level changes were determined. The data collected from the experiment

were used to assess the uncertainty of Sy using Moench's, Neuman's, and Theis' models, where the Markov chain Monte Carlo method was used to sample the data. The results showed that most of the uncertainty in this method is due to uncertainty in the parameters.

Another study by Crosbie et al. (2019) treated the specific yield as an unmeasurable variable and attempted to assess the uncertainty by applying a rejection sampling approach using probabilistic estimates of net recharge from two methods. This study was applied in 4 different catchments, and the results showed that constraining the specific yield in the recharge estimate provides consistent output.

The parameter uncertainty related to base flow separation conducted using a two-parameter recursive digital filter (TPRDF) was investigated by He et al. (2022). In this study, the uncertainties due to two parameters, namely, the recession constant and maximum base flow index (BFI), were investigated. This study revealed that optimized values of BFI and time-scaled recession constants can be used to reduce the uncertainty in base flow separation (He et al., 2022). The sensitivity of the BFI was investigated by Yang et al. (2019) using base flow separation by a two-component hydrograph separation method with conductivity as a tracer. The uncertainty in the BFI can be reduced by half when considering the mutual offset of the measurement errors in conductivity and stream flow(Yang et al., 2019).

Model structure uncertainty: The water table fluctuation (WTF) method of recharge estimation has been shown to be a very reliable estimation method compared to other methods (Delottier et al., 2018, Heppner and Nimmo, 2005). The method assumes that recharge to an unconfined aquifer is proportional to the change in the water level in the saturated zone (Scanlon et al., 2002, Healy, 2010). This method is related to the volume of fillable porosity or drainable porosity (Park, 2012). Of course, falling water table analysis should be treated with drainable porosity as the storage parameter, while fillable porosity should be used for WTF analysis of rising water tables. The simplified mathematical representation of this method (Eqn. 2) is as follows (Park, 2012):

$$\boldsymbol{R = Sy\frac{dh}{dt}} \text{----------Eqn. 2}$$

where Sy is the specific yield and $^{dh}/_{dt}$ is the change in the groundwater table over time.

In the above simple model, the effects of delayed recharge flux, unsaturated drainage and other important futures of groundwater table level time series are not well represented (Park, 2012). The delay from the rainfall event to the water table response and the delayed gravitational flow from the unsaturated zone to the saturated zone can cause an undetectable recharge rate (Jeong and Park, 2017). Another controversial concept of time-variant specific yield was also tested using field observations as a model uncertainty reduction measure by Crosbie et al. (2005). When using the WTF method, other sources of uncertainty are

air entrapment and water table recession due to the lateral flow of water (Walker et al., 2019). One way of reducing such problems is by using an effective water table rise (Delottier et al., 2018, Cuthbert, 2010).

The base flow separation method can only be applied to gain streams (Rutledge, 2007). Typically, to identify the loosing or gaining section of a river, additional modelling or assessment is needed. There is also a need for this method even when following a reliable separation technique. At some reaches of the stream bank, when the water level is high enough, water is stored in the bank and discharged when it is low, which cannot be considered recharge (Walker et al., 2019). The uncertainty level will also increase if we do not include extraction, evapotranspiration, or underflow to deep aquifers (Scanlon et al., 2002). Scanlon et al. (2002) stated that the reliability of this method depends entirely on the validity of the assumptions and good estimation of the source.

One of the most widely applied methods for separation is the recession curve displacement method. The watershed for which this approach is applicable ideally has uniform rainfall and geological conditions (Healy, 2010). Additionally, this method cannot be applied for watersheds containing flood control structures (Rutledge, 2007). Other methods of separation, including the one described above and their corresponding structural uncertainties, are explained in detail by Healy (2010) Heppner and Nimmo (2005).

(Gallart et al., 2007) applied the generalized likelihood uncertainty estimator (GLUE) uncertainty assessment tool to investigate the impact of understanding and tailoring catchment information to reduce both surface water and base flow simulation uncertainty. The study was conducted using TOPMODEL (Beven, 1997). The findings of the study confirmed that the uncertainty level can be reduced by introducing conditioned water table records and the distribution of parameters obtained from point observations (Gallart et al., 2007).

2.1.4. Regional-scale recharge uncertainty

Recharge uncertainty assessment of mesoscale watersheds requires in-depth knowledge of the hydrogeological processes of the focal area. In developing countries aided by a lack of sufficient ground observations, it has been a challenge to investigate/conduct recharge uncertainty assessments. This subsection aims to highlight the impact of data scarcity on recharge uncertainty despite differences in the hydrometeorological characteristics of the selected watersheds. Contrasting watersheds where multiple recharge estimations are attempted using different methods are selected to illustrate the impact of data scarcity. Generally, as discussed in the introduction section, very little has been done to investigate groundwater recharge uncertainty in mesoscale watersheds of Ethiopia. A very good summary of studies on groundwater recharge conducted in Ethiopia was presented by Walker et al. (2019). In this study, 9

methods of estimation were implemented to provide insight into uncertainty characterization in three catchments in the Blue Nile River basin (Walker et al., 2019). This study aimed to assess the discrepancies caused by a lack of recharge interpretation and uncertainties due to the inherent assumptions of a specific method. The uncertainty result was presented with a higher and lower limit of actual recharge, and it was noted that the type of recharge should be inferred for each method.

The uncertainty in most studies is presented by implementing multiple methods at a site and providing a range of recharges. However, as noted by Walker et al. (2019), in addition to the inherent assumptions and recharge type estimated by each method, the data-driven method selection procedure followed by uncertainty analysis results in incomparable outputs. In the upper Awash basin, the implementation of four methods (of which three estimated actual recharge while the remaining are potential and minimum recharge) has resulted in incomparable outcomes (Berehanu et al., 2017). Similarly, the study highlighted that the results obtained cannot be used as an uncertainty range descriptor. Such an approach, when implemented properly, can provide good insight into the range of recharge uncertainty but cannot identify the uncertainty in either individual methods or input data.

Inevitable uncertainties in estimation can be illustrated by observing multiple estimations (Table 1) in a particular region. Below are tabulated estimations of different regions by different methods and multiple scholars. An elaborate discussion on the causes and discrepancies of the results is given as follows by selecting one particular watershed (i.e., the Hawassa Lake watershed, Ethiopia). The main similarity between the below tabulated multiple estimation for similar watersheds using different methods (Table 1) is that all locations suffer from data scarcity and the lack of adequate available data for accurate recharge estimation.

A review of previous studies focused on assessing the recharge of the Densu River basin indicates that the average annual recharge reaches more than 10% of the annual average rainfall (Akurugu et al., 2022). However, Akurugu et al. (2022) reported that a similar basin water balance approach yielded a smaller magnitude of mean annual recharge for the same watershed. These findings illustrate that recharge uncertainties increase where field observations are scarce. Furthermore, the importance of coupling different recharge estimation methods for data-scarce regions was illustrated by Adomako et al. (2010). Furthermore, discrepancies between the estimates obtained using different methods for similar regions highlight the importance of appropriately selecting methods for specific objectives. A similar deduction can be drawn by observing the discrepancies in the Verlorenvlei catchment recharge estimation.

Table 1 Variability in recharge estimates for similar study areas using multiple methods

Publication	Recharge type	Method	Recharge (mm/year)	Annual precipitation (mm/year) (year)	Basin	Remark	Country
(Zemedagegnehu, 2020)	Actual	BWB	211.5	1039.35 mm (1991-2010)		Basin average	
(Kebede, 2013)	minimum	BFS	50-150	NA		Basin average	
(Lin, 2020)	Potential	NM	138.8	671.50 mm 2006-2016		Basin average	
(Ayenew and Tilahun, 2008)	Actual	SWB & BWB	17.9-80.3	1030 mm (1970-2004)	Lake Hawassa	For 200 m resolution (variable recharge)	Ethiopia
(Yeneneh, 2014)	Potential	NM	19.63% of AP	818-1117 mm (1969-2013)		Basin average	
(Lemlem, 2008)	minimum	BFS	90-190	1038 mm (1970-2004)		Basin average	
(Adomako et al., 2010)	Potential average Actual	NM IPS BWB	182 110 135	1310 mm (2006-2008)	Densu River @ sample site Ayikae Doblo	Isotope concentration (O18 and H2) were sampled from soil depth profile and rainfall for both seasonal peak shift and numerical modelling, whereas nearby measurements where used for calculating simplified water balance	Ghana
(Adomako et al., 2010)	Potential Average Actual	NM IPS BWB	94 120 75	1310 mm (2006-2008)	Densu River @ sample site Teacher mante		
(Adomako et al., 2010)	Potential Average Actual	NM IPS BWB	101 250 21	1737 mm (2006-2008)	Densu River @ sample site Adwumuko		
(Akurugu et al., 2022)	Actual Actual	BWB BWB	120-153 172.2	1700 mm (NA)	Densu River	WB for different geological formations Basin Average	
(Berehanu et al., 2017)	Actual Actual Potential Minimum	BWB CMB NM BFS	131 135 157 91.25	1170 mm (NA)	Upper Awash	Basin average by considering subsurface inflow Basin average Basin Average using HYDRUS-1D Basin Average	Ethiopia
(Tolera and Chung, 2021)	Actual	BWB	179	1207 mm (1988-2004)		Basin Average using SWAT	

(Eilers, 2018)	Actual	CMB	20-27 40-53		Upper valley sub catchment Sub catchments of Mountain ranges
(Watson et al., 2020)	Actual	CMB	4.2-5.6% of MAP 11.4-15.1% of MAP	Verlorenvlei catchment	Upper valley sub catchment SA Sub catchments of Mountain ranges

Note: BWB, basin water balance method; BSF, base flow separation method; SWB, soil moisture balance method; AP, annual precipitation; MAP, mean annual precipitation; CMB, chloride mass balance; NM, numerical modelling; IPS, isotope peak shift; SA, South Africa.

Over the years, many studies have been conducted on the surface and subsurface hydrological processes of the Lake Hawassa watershed. The results of each study have significant implications for basin recharge. A significant discrepancy can be observed in previous studies (Table 1) on recharge values in the basin. Among the factors affecting recharge, the spatial and temporal distributions of land use/cover have been assessed in the Hawassa watershed. The implications of temporal land use dynamics for the environment (Degife et al., 2019), groundwater recharge (Lemlem, 2008), catchment hydrology (Orkodjo, 2014), and soil erosion (Degen, 2016) are some of the areas of interest. The common agreement of these studies is that forest cover, shrub land cover, and wet land have drastically decreased, while cover, such as urban cover, bare land cover, and agriculture, has increased (Degife et al., 2019, Degen, 2016), which implies that there has been significant change in hydrological processes.

However, the meteorological data analysis revealed decreasing and constant trends in evapotranspiration and temperature, respectively (Lemlem, 2008, Orkodjo, 2014). These results contradict the inference that recharging should occur. Land use/land cover studies (Eshete, 2009, Degen, 2016, Degife et al., 2019) in the basin have highlighted that there is a significant change from forest/shrub land to bare land and agricultural land. Such a change results in an increase in recharge (Pan et al., 2011, Owuor et al., 2016); in contrast, a study by Lemlem (2008) suggested a declining recharge trend over 3 decades (i.e., until 2000) in the Hawassa Lake catchment. Sources of uncertainty can be accredited to input data as well as parameter uncertainty and model structural uncertainty.

Aquifers with higher permeability recharges are highly sensitive to the amount of rainfall (Xie et al., 2018). Recent fluctuations in the lake level of Hawassa have been related to the impacts of El Nino and La Nina (Belete et al., 2017, Belete et al., 2016). Others relate the scenario to a significant interaction of overland flow and groundwater with lake level fluctuations (Ayenew and Tilahun, 2008, Ayenew and Becht, 2008). The preceding argument is challenged by scarce meteorological data records (Belete et al., 2017), while the second statement is vulnerable to the limited spatial distribution of other input data and modelling uncertainty (Ayenew and Tilahun, 2008). From both studies, one can infer that significant recharge causes lake level fluctuations or that lake level fluctuations are assisted by increased surface runoff (Tessema, 2004) and recharge. However, there is a gap in the identification of the contributions of surface runoff and subsurface processes to these fluctuations.

Environmental stable isotope studies are being used to better understand the hydrological and hydrogeological processes in Ethiopia (Tekleab et al., 2014, Tolke and Ayenew, 2019). In one of the main Ethiopian Rift (MER) subbasins (i.e., the Abaya-Chamo Basin) Tolke and Ayenew (2019), environmental isotope and geochemical analyses were implemented to characterize the recharge source, type, and hydrogeological system of the basin. This study provides insight into the recharge mechanism and subsurface flow of the basin. A significant contribution to creating a better understanding of the hydrogeology of Lake Hawassa was presented by Tessema (2004), who used isotopic and geochemical investigations. This study concluded that the recharge elevation for the basin was 2000 m.a.s.l. The lake receives inflow from the eastern, western, and southern shorelines, while it is lost from the northern shoreline. As far as this review is concerned, no research has investigated recharge uncertainty using isotopic analysis in the basin.

2.2. Uncertainty assessment tools

Uncertainty has been a challenging task in many recharge estimation efforts by different researchers. Concepts such as unpredictability, imprecision, and variability are used to describe uncertainty in modelling (Contreras et al., 2018). Unpredictability can be due to natural trends in inputs and errors in observations, whereas the inaccuracy of boundary conditions, model simplification through parameter reduction, and other similar sources of uncertainty are related to imprecision (Moges et al., 2021). Spatial and temporal uncertainties are usually best described with the variability concept and are usually difficult to quantify (Contreras et al., 2018). Model initialization can also be a source of uncertainty (Yu et al., 2019). Uncertainty sources such as parameters, structure, and inputs are included in some of the sources of uncertainty discussed above. Further discussion on the reviewed literature focusing only on the parameters,

model structure, and input uncertainty evaluation tool is presented as follows. The objective of this section is not to discuss available methods or to repeat previous reviews but rather to highlight available methods.

Recent developments in the field of uncertainty analysis include the implementation of deep learning algorithms. The parameter uncertainty assessment using Bayesian long short-term memory (LSTM) through a Bayesian sampling approach called stochastic variational inference has been proven to perform better than the traditional Bayesian linear regression model (Li et al., 2021). Another deep learning tool for uncertainty assessment is Monte Carlo dropout with an input-dependent data noise term (MCD+N)(Gal and Ghahramani, 2016). Fang et al. (2020) implemented MCD+N to investigate the parameter uncertainty in soil moisture prediction.

Input uncertainty: Simple water balance and complex modelling methods require meteorological, hydrological, topographic, geological, and other types of input data. However, very few methods are available to explicitly quantify the uncertainty from such inputs (Ajami et al., 2005). A method introduced by Huard and Mailhot (2006), as well as methods such as BATEA (Kavetski et al., 2003) and IBUNE (Ajami et al., 2005), can explicitly assess input uncertainty (Balin et al., 2010). Most input uncertainty studies have focused on precipitation uncertainty assessments (Moges et al., 2021). Nevertheless, when considering subsurface hydrological processes, other inputs, such as land use/cover and evapotranspiration, need to be considered. Scarce calibration data are also another challenge when assessing such uncertainty (Muñoz et al., 2014). Simple multiplicative error models have also been studied to decrease the uncertainty in mean aerial rainfall (McMillan et al., 2011, Balin et al., 2010). McMillan et al. (2011) highlighted that such a simple method is better at capturing the tail distribution than the lognormal multiplier distribution. However, approaches such as this are limited in capturing the 'true' rainfall distribution (Moges et al., 2021).

The generalized likelihood uncertainty estimator (GLUE) is also used to evaluate the uncertainty in point rainfall observations (Muñoz et al., 2014). The GLUE approach results in uncertainty related to the input combined with parameter uncertainty (Muñoz et al., 2014, Moges et al., 2021). Xie et al. (2018) applied a three stage procedure to analyse the uncertainty due to the input data, (which was adopted for this study). After defining the different recharge zones the first step was to generate possible parameter combinations, followed by identifying acceptable model realisations, and finally analyse the long term recharge statistics for each recharge zone. The uncertainty of recharge for 13 model setups was presented by producing spatial CV of the long term mean annual recharge produced by each accepted model realisations.

Parameter uncertainty: Parameter uncertainty analysis has been given significant attention and has benefited from multiple assessment methods (Moges et al., 2021). Uncertainty analysis in recharge

estimation methods has also been the main objective of many studies (Delottier et al., 2018, Crosbie et al., 2019). The shared approach in such analysis is that model realization is created by applying a Monte Carlo simulation on a bounded parameter range (based on prior knowledge), followed by an investigation of the results from a specified objective function. Tools such as GLUE and the model-independent parameter estimation and uncertainty analysis tool PEST (Doherty et al., 2010) mainly rely on such approaches, while others such as DREAM rely on Bayesian statistics (Moges et al., 2021).

Model structure uncertainty: Usually, model structural uncertainty arises due to simplification of natural processes (Ajami et al., 2005). To this end, very good prior knowledge of the existing conditions is an appropriate way to decrease model structure uncertainty (Duan et al., 2007). Uncertainty assessment tools such as IBUNE are equipped with a model averaging approach (Bayesian model averaging) to reduce structural uncertainty (Ajami et al., 2005). For a detailed discussion of model averaging tools, refer to the reviews by (Moges et al., 2021, Fragoso et al., 2018).

2.3. Discussion

***Discussion of identified gaps*:** Notably, each method has potential applicability and strength over the other methods. Tracer methods have good quality in terms of acquiring subsurface processes and information that cannot be acquired by other methods (Vitvar et al., 2005). Some methods provide potential recharge, while others provide actual recharge to the water table. Methods such as the CMB and BWB methods are used to determine actual recharge, while the SWB and WTF methods can be used to determine potential groundwater recharge for a given aquifer. Readers are encouraged to refer to Walker et al. (2019) and Lorentz et al. (2003) for more details on the types of recharge for different recharge determination techniques. Using multiple methods is recommended for a reliable estimation of recharge (Scanlon et al., 2002). Additionally, research involving the use of tracer methods coupled with base flow separation methods has shown good results (Stewart et al., 2007). Previous knowledge of watershed and aquifer properties provides additional useful information for qualitative uncertainty analysis (Xie et al., 2018). It is also very important to understand the assumptions of a method applied to a specific catchment.

Tracer-based methods are reported to suffer from the inability to capture the contribution of direct recharges in faulty (i.e., geological cracks and fissures) regions (Berehanu et al., 2017). Another limitation is the gap in spatial representation of sampling for the CMB method (Crosbie et al., 2018). However, such methods are advantageous for establishing benchmark point measurements over other preferred methods for spatial recharge estimation, such as numerical modelling. Modelling efforts have disadvantages in structuring, as most models emphasize either surface hydrology over groundwater hydrology or vice versa (Bear et al.,

2010) (Dile et al., 2018). Hence, coupled models that give equal emphasis to surface and groundwater hydrology are recommended for future studies.

Several studies in different parts of the world conducted using two or more methods reported different magnitudes of recharge (Walker et al., 2019). These discrepancies are attributed to the individual assumptions of each method as well as the spatial and temporal resolutions of the estimations (Scanlon et al., 2002). This means that there are inherent uncertainties in each estimation. Generally, the uncertainty in each method arises from the uncertainty of the parameter, model structure, and input data.

The uncertainty analysis of this important hydrogeological parameter spans from simple range determination (i.e., maximum recharge and minimum recharge) to input parameter uncertainty analysis. There are also significant studies that have been conducted to quantify the spatial uncertainty of estimations (Alcalá and Custodio, 2014, Crosbie et al., 2018). Considering the importance of mapping recharge for large areas, spatial uncertainty needs to be investigated further. In general, the main findings of this review on recharge uncertainty are as follows.

Conceptualization and good prior knowledge of aquifers as well as watersheds are important before conducting uncertainty analysis (Moges et al., 2021, Xie et al., 2018). Considering the differences in the process representations and assumptions of different models, the selection of an adequate model requires prior knowledge (Gupta et al., 2012). Otherwise, misrepresentation/oversimplification of any existing process will result in structural uncertainty (Moges et al., 2021). Somaratne and Smettem (2014) demonstrated the inadequacy of the conventional CMB recharge estimation method for watersheds with deep faults. Additionally, parameter uncertainty analysis based on Monte Carlo simulation requires bounded parameter values gained from prior knowledge of the existing conditions (Xie et al., 2018). Furthermore, prior knowledge is essential for minimizing uncertainties emanating from boundary conditions and model initialization.

Therefore, considering the variability in estimation methods and watershed characteristics, the inherent assumptions of available methods need to be further assessed for uncertainty (i.e., model structure uncertainty). Structural uncertainty can potentially be reduced by using multiple methods of estimation (Scanlon et al., 2002, Berehanu et al., 2017). Furthermore, multimodel ensemble and averaging techniques should be investigated in depth. Mustafa et al. (2020) introduced an integrated Bayesian multimodel uncertainty estimation framework for analysing uncertainties emanating from alternative conceptual models. Multimodel averaging approaches result in far less uncertainty than single model estimates (Moges

et al., 2018). Despite the inability to consider all possible model structures, coupling multiple models can be used to assess the uncertainty in model structure (Moges et al., 2021).

In data-scarce areas, the spatial uncertainty of recharge is mostly due to the uncertainty of the input data (Crosbie et al., 2018, Alcalá and Custodio, 2014, Westerhoff et al., 2018). Zhang et al. (2018a) illustrated the influence of spatial rainfall uncertainty on hydrological simulations using the bootstrap method. Similarly, rainfall and geology input data have been observed to cause more uncertainty in spatial recharge estimates Westerhoff et al. (2018). Other input data, such as evapotranspiration and soil moisture, were investigated by Xie et al. (2018). The results indicate that input spatial uncertainty could be propagated to simulation outputs (Montanari and Di Baldassarre, 2013). Therefore, further investigations are required to fully understand the uncertainty from input data for different hydrological settings and input data.

Future research directions: Discrepancies between recharge estimation using multiple methods at a single location (Table 1) indicate that more research should be done to investigate the root cause of the variation in recharge estimation. The uncertainty analysis of input variables, especially physical methods, needs more work since such works will be useful in guiding modellers on which parameter to focus on. As discussed earlier, recharge is a spatially variable phenomenon and should be treated as such. This implies that future work on uncertainty analysis of scaling-up techniques is imperative. In general, from the findings of this review, the following areas are recommended for future studies.

Uncertainties in physical-based recharge estimation are one of the identified focus areas for future research. This review indicates that little has been done to investigate the uncertainty of input data for physical-based methods. A physical-based recharge estimation method for WTFs requires water table depth time series data as an input (Heppner et al., 2007, Delin et al., 2007, Heppner and Nimmo, 2005). Despite advancements in water table depth measuring equipment, continuous calibration of devices is required and lacking in most developing countries (Dile et al., 2018). Hence, further investigations are needed to quantify the errors induced by such unreliable input data. Furthermore, the spatial representativeness of observation wells in light of spatial error propagation and distribution needs to be investigated further.

Alternatively, current advancements in input data assimilation from multiple sources have proven to reduce errors efficiently for further application (Gebremedhin et al., 2021). Implementing a merging technique using two or more rainfall data sources enables the capture of the complementary merits of each source, which could reduce bias in satellite rainfall products (Jongjin et al., 2016). Hence, the assimilation/fusion of different sources of meteorological and hydrological input data for spatial uncertainty analysis should be considered in future studies. Similarly, more reliable estimates of groundwater recharge can be obtained by jointly applying two methods (Gumuła-Kawęcka et al., 2022). Gumuła-Kawęcka et al. (2022) illustrated

the potential to decrease and understand uncertainty by combining two methods using the WTF method, which is susceptible to errors from both the RISE and master recession curve (MRC) approaches, and the HYDRUS-1D (Simunek et al., 2005) numerical model. This was accomplished by separating the rise of the water table due to recharge from the rise and/or fall of the water table level due to outflows to and/or inflow from nearby surface water bodies, which corrects the misassumption that all rises are caused by recharge in the WTF method. Therefore, future works focusing on coupling methods to better understand the uncertainties in recharge assessments should be considered.

Among the available methods, one widely used method of estimation is modelling the basin water balance, which is reported to yield rough estimates (Berehanu et al., 2017, Adomako et al., 2010). The analysis of the uncertainty of input data to determine the uncertainty of outputs is a never-ending approach as the efficiency of the input data acquisition method progresses. For a more efficient hydrological analysis, existing scarce meteorological data in the Hawassa watershed need to be upgraded using satellite rainfall products. However, from the review of previous studies in the basin, it is clear that there is a gap in quantifying the uncertainty in estimating recharge. Additionally, the need for stipulating uncertainty from scarce rainfall records is observed as a gap, i.e., for recharge estimation. Furthermore, studies have proven that the other major input meteorological data with a significant contribution to recharge uncertainty is potential evapotranspiration. The significance of ET estimates is significant for watersheds such as Lake Hawassa, where other meteorological measurements essential for the estimation/determination of potential evapotranspiration are highly scarce and most estimates rely on simple temperature methods or depend on measurements conducted outside the watershed boundary. Hence, the assessment of alternative ET data sources with related groundwater recharge estimate uncertainty is needed.

CHAPTER THREE

3.0. Methodology

3.1. Study Area Description

Location and Topography

The Tikur-Wuha watershed covers an area of 670 km^2. It is the only perennial river flowing into Lake Hawassa. Steep mountain ranges cover the northeastern and southeastern parts of the watershed, while gently sloping land covers the northwestern to southwestern parts of the watershed. Nine perennial streams feed the wetland located at the center of the watershed, and the Tikur-Wuha River drains to Lake Hawassa through that wetland (Figure 2).

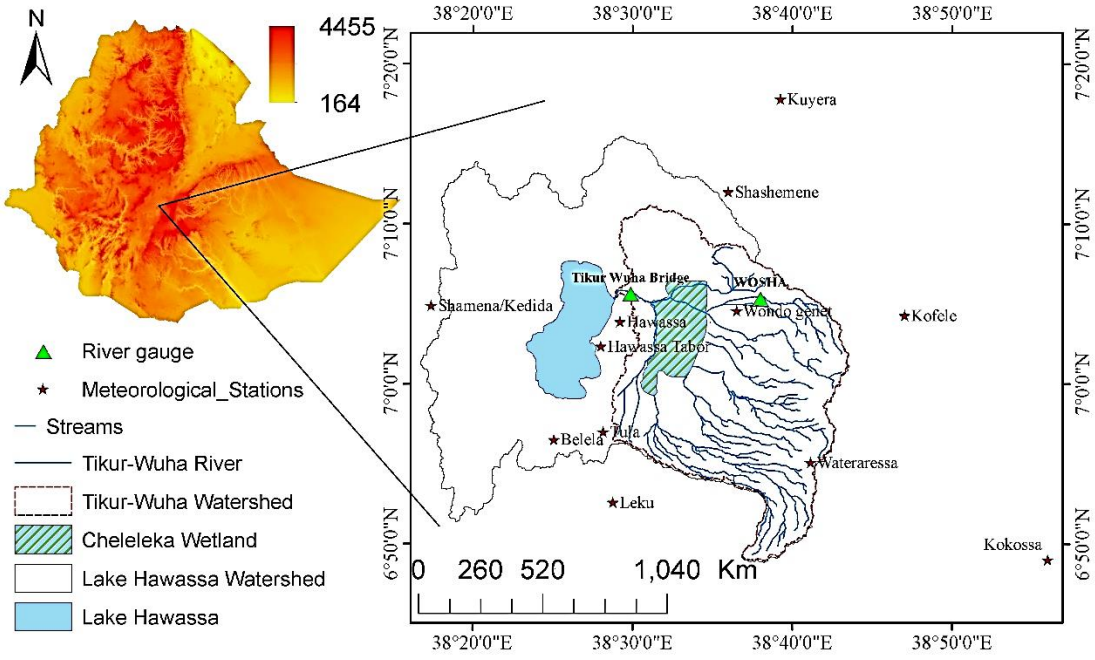


Figure 2 Location map of the study area

The Tikur-Wuha watershed is located at 60°48’00’’–70°10’00’’N latitude and 38°26’30’’–38°43’00’’E longitude. It is at an elevation between 1643 and 2976 m above mean sea level. According to Mohammed et al. (2020), 54% and 21.6% of the Tikur-Wuha watershed has slopes ranging from 8% to 30% and 3% to 8%, respectively, while 3% and 19% of the watershed have slopes < 3% and > 30%, respectively (Mohammed et al., 2020).

Climate

The Tikur-Wuha catchment has a tropical subhumid climate (Moist Weyna Dega). The mean annual rainfall ranges from 954.5 mm in the western and northwestern parts (lowlands) to 1152.5 mm in the southeastern part (mountainous section of the catchment). As reported by Belete et al. (2017), up to 70% of the annual rainfall in Ethiopia occurs when the intertropical convergence zone (ITCZ) lies over northern and southern Ethiopia from June to September (Belete et al., 2017). The seasonal rainfall cycle, characterized as a bimodal cycle (i.e. the watershed receives two wet seasons) (Abraham and Muluneh, 2022b), is mainly controlled by the movement of the ITCZ (Belete et al., 2017).

The seasonal rainfall in the watershed can be quantified as follows: Thirty percent of the annual rainfall occurs between March and May, with the seasons from June to September and October to February contributing 50% and 20% of the rainfall, respectively (Belete et al., 2017). The average maximum and minimum temperatures of the watershed range from 11 to 20°C, and the hottest dry month of the watershed is March (Abraham and Muluneh, 2022a).

Hydrology

The Tikur-Wuha subwatershed is located within the geographically closed watershed of Lake Hawassa, where uncontrolled irrigation practices are increasing. The Tikur-Wuha River feeds Lake Hawassa, and the water from the Tikur-Wuha watershed meets the water requirements of households, livestock, irrigation, industries, and the environment. Nine streams, namely, Bele-Lango, Wesha, Werka, Hallo, Shenkora, Gomesho, Wedesa, Abosa, and Gelchacha, flow through the Tikur-Wuha watershed, which drains to the Cheleleka wetland (Kebede et al., 2014). The aggravated land cover changes have resulted in the disappearance of one lake within the Cheleleka wetland (Wondrade et al., 2014). The Cheleleka wetland exhibits flow stabilizing properties by serving as an intermediary between the upper catchments of the nine rivers, where the water flow is rapid, and the Tikur-Wuha River, where the water flow is slow. The wetland receives water from the nine streams and channels to one outlet to feed the Tikur-Wuha River, which then drains into Lake Hawassa. The watershed outlet is located at the Tikur-Wuha Bridge. The watershed covers an area of 645.75 km^2, and the estimated average annual flow of the river for the period between 1987 and 1996 was 132.6 mm, calculated by dividing the average annual stream flow volume by the total area of the watershed (Abraham and Muluneh, 2022b, Abraham et al., 2022).

Land use and soil

Studies focusing on land use/cover classifications have shown that since the 2000 Gregorian Calendar (G.C.), most watersheds have become croplands used for small-scale plantations, rain-fed crop fields, and

perennial cash crops. According to Wondrade et al. (2014), between 1973 and 2011, cropland coverage increased by 12.8% (Wondrade et al., 2014). Other major types of land use/cover in the watershed include grasslands, wetlands, forests, urban areas, and shrublands. Most of the forest cover is in escarpments, and the rift floor is covered with wetland, grasslands, and cropland. A significant change in land use/cover dynamics is the disappearance of Lake Cheleleka, which covers 1300 ha (Wondrade et al., 2014). The disappearance of Lake Cheleleka can be attributed to environmental degradation due to extensive deforestation, climate change, and other anthropogenic stressors (Wondrade et al., 2014). However, no studies have been conducted to investigate in particular the real causes that led to the disappearance of the lake. In general, 73.4% of the watershed experienced land use/cover changes between 1972 and 2017 (Degife et al., 2019).

The soil map of the watershed (Figure 5) indicates that sand, clay loam, sandy loam, and silt cover 17.24%, 24.78%, 51.01%, and 17.10% of the watershed, respectively. According to Mohammed et al. (2020), Leptosol, Vertic cambisol, Vertic Luvisol, Vitric Andosol, Haplic Luvisol, and Eutric Cambisol are the major types of soil found in the Tikur-Wuha watershed, with Vertic Luvisol and Haplic Luvisol dominating 39.6% and 39.2% of the watershed, respectively (Mohammed et al., 2020).

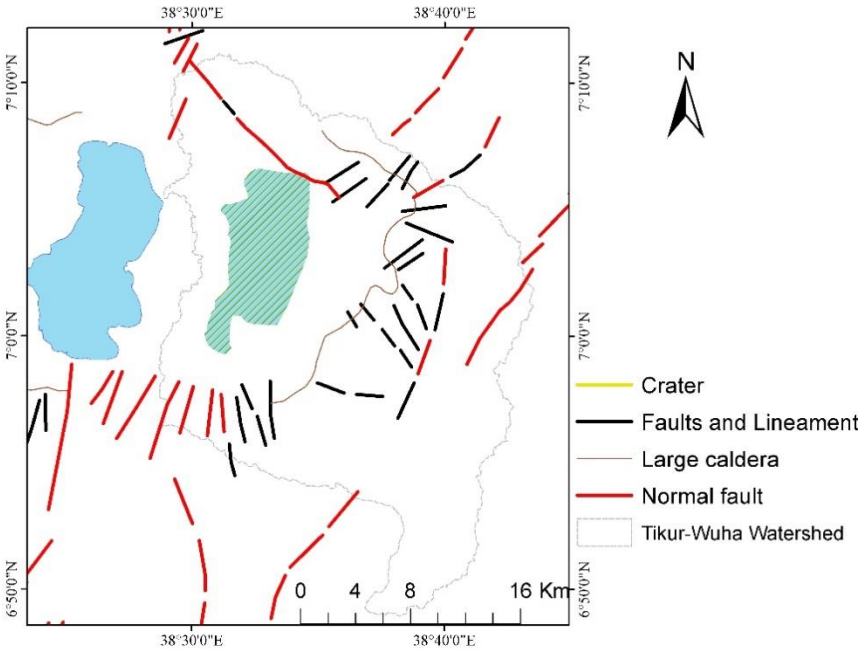


Figure 3 Structural geology of the study area

The geological structure map of the study area (Figure 3) shows the presence of densely populated normal faults in the southeastern part of the watershed. The drainage pattern of the study area (Figure 2) overlaid on its geological structure map (i.e., showing fault lines in Figure 3) indicates that most of the fault lines

intersect the main drainage lines, which has the potential to increase surface flow interception through point recharges. Ayenew and Tilahun (2008) argued that simulations conducted for watersheds exhibiting densely populated fault structures should consider fractured volcanic aquifers as porous media, enabling heavily fractured regions to exhibit high hydraulic conductivities. Their study also highlighted that the existing fault structures and newly emerging cracks in the Lake Hawassa basin have the potential to create point recharges.

Highly permeable and productive lacustrine deposits to medium permeable acidic volcanic, and least permeable obsidian flow units of geological formation characterises the aquifer system of Lake Hawassa watershed (Ayenew and Tilahun, 2008). Additionally, the caldera flow composed of alluvial deposits overlay with thick lacustrine sediments is characterised as moderately productive and permeable aquifer, whereas in the eastern highland containing mostly welded tuff exhibit low-to-moderate potential aquifers (Ayenew and Tilahun, 2008). Readers are encouraged to read Ayenew and Tilahun (2008) further for detailed description and aquifer characterisation of Lake Hawassa watershed.

3.2. Dataset

The inputs to the selected methods (Table 2) are three base maps (i.e., land-use, soil, and digital elevation model (DEM)) and hydrometeorological data. As one of the specific objectives of this study was to assess the recharge estimation uncertainties associated with input rainfall data, two sets of rainfall records were used. The first rainfall record is point observations from rain gauges in and around the watershed, whereas the second dataset is a merged rainfall dataset from the Climate Hazards Group Infrared Precipitation (CHIRP) satellite rainfall product with available ground station records. Furthermore, this study aimed to investigate the impact of input evapotranspiration (ET) on the uncertainty of recharge; hence, in addition to the estimates of ET from ground observations of meteorological data, satellite-based reanalysis ET data called the Global Land Evaporation Amsterdam Model (GLEAM) were applied.

Meteorological Data

Available meteorological parameters were gathered for stations in and around the Lake Hawassa watershed. Accordingly, daily meteorological records from two synoptic meteorological stations, two ordinary meteorological stations and eight rain gauges were obtained from the National Ethiopian Meteorological Agency (EMA). Synoptic meteorological stations measure precipitation, wind speed, sunshine hours, maximum temperature, minimum temperature, and relative humidity. On the other hand, ordinary stations only record precipitation and temperature (i.e., maximum and minimum), while the remaining stations only provide precipitation records.

Table 2 Input datasets and their respective data sources

Data Description	Source	Detail	Time scale
Ground Meteorological time series data	National Ethiopian Meteorological Agency	2005-2018	daily
CHIRPS satellite rainfall estimates (SRE)	https://data.chc.ucsb.edu	2005-2018	daily
CHIRP satellite rainfall estimates	https://data.chc.ucsb.edu	2005-2018	daily
GLEAM Ep_3.8a	http://www.gleam.eu/	2005-2014	daily
Digital elevation model	http://gdex.cr.usgs.gov/gdex/	30 m resolution	NA
Soil map	Geological Survey of Ethiopian	-	
Daily stream flow data	Mistry of water and energy of Ethiopian	At Wosha from 2005 to 2014	daily
		At Dato village from 2005 to 2014	daily

The ground-observed rainfall data were obtained from the National Ethiopian Meteorological Agency (EMA), whereas the Climate Hazards Group Infrared Precipitation (CHIRP) satellite rainfall product and its modified version with ground station data, CHIRPS 2.0, for the time period from January 2008 to August 2018 on a daily time scale were obtained from https://data.chc.ucsb.edu. However, it should be noted that ground observations are available for different periods since the establishment of each station and that the specified length of the period was selected to ensure the maximum overlapping length of observations from available stations with minimum missing data (on average, up to 20%). In the study watershed, snow formation is less likely to occur; hence, the satellite precipitation product is referred to as the satellite rainfall product (SRP).

Freely available CHIRPS version 2.0 and CHIRP SRP data, with a resolution of $0.^{05^\circ}$ for monthly, decadal, pentadal and daily time scales (Funk et al., 2015), were downloaded from https://data.chc.ucsb.edu at a daily time scale. The climate hazards group infrared rainfall with station dataset (CHIRPS) is a quasiglobal product dependent on an algorithm that fuses estimates from infrared cold cloud duration (CCD) observations and ground observations (Funk et al., 2015). In the preparation of the rainfall dataset, all records from 11 stations were used to create the merged rainfall data for the second objective of the study (Beyene et al., 2023). However, it should be noted that based on the available records, five stations were

used for the year 2005, while eight stations were used for the year 2006. For the conventional method of areal rainfall estimation for the Thiessen polygon, only four stations in and around the Tikur-Wuha watershed were used for the second objective of the study.

According to the station location information gathered from the EMA, only six stations are found within the watershed. To complement missing data and represent the western part of the watershed as opposed to the densely populated station distribution in the eastern part, an additional six stations surrounding the watershed were considered. Further analysis of the missing data rate and seasonal contribution (Table 3) was conducted to indicate the status of the collected data. To enhance the accuracy of the filling gauge rainfall records, a total of 28 stations (Figure 4) were used, including those used for this study. Furthermore, whenever needed, the missing data were filled using the arithmetic mean method.

Table 3 Assessment of missing rainfall records and rainfall contributions during dry (November to March) and wet (April to October) periods

Station	Station Code	Elevation (m)	Rainfall Percentage		Missing Data	
			Wet period (%)	Dry period (%)	Wet period (%)	Dry period (%)
Kuyera	KU	1936	85.2	14.8	10.5	11.1
Belela	BE	1866	88.5	11.5	7.9	16.3
Leku	LE	1879	80.8	19.2	6.5	7.1
Kokossa	KOK	2400	77.3	22.7	32.8	36.5
Shamena_Kedida	SHK	2072	82.1	17.9	6.7	11.2
Wondo_genet	WO	1880	81.6	18.4	7.8	1.7
Tula	TU	1873	78.1	21.9	4.0	3.5
Hawassa	HA	1750	83.7	16.3	0.1	3.1
Hawassa_Tabor	HW	1750	82.8	17.2	5.2	1.9
Wateraressa	WT	2631	83.3	16.7	16.1	19.8
Shashemene	SHA	1927	80.7	19.3	17.0	23.7
Kofele	KOF	2620	78.1	21.9	5.4	6.7

Data quality assessments of the outlier Grubbs and Beck (1972) test (G-B), homogeneity and stationary Mann–Whitney test (1947), independence and stationary test Wald-Wolfowitz (1943), and consistency test (double mass curve) were conducted on the meteorological data records. Accordingly, corrections for consistency were made to the rainfall records of the Wateraresa station, where significant deviation was observed and slight deviation was observed in the consistency check for the Hawassa and Shashemene stations. Other quality tests were conducted for sample data (i.e., daily maximum and annual records) from each station. Tests for independence, homogeneity and stationarity indicated that the data from all the stations were homogeneous and independent. Furthermore, based on the assessment of rainfall record data

quality, records from the Shashemene station are excluded from further use in the modelling process of the second objective.

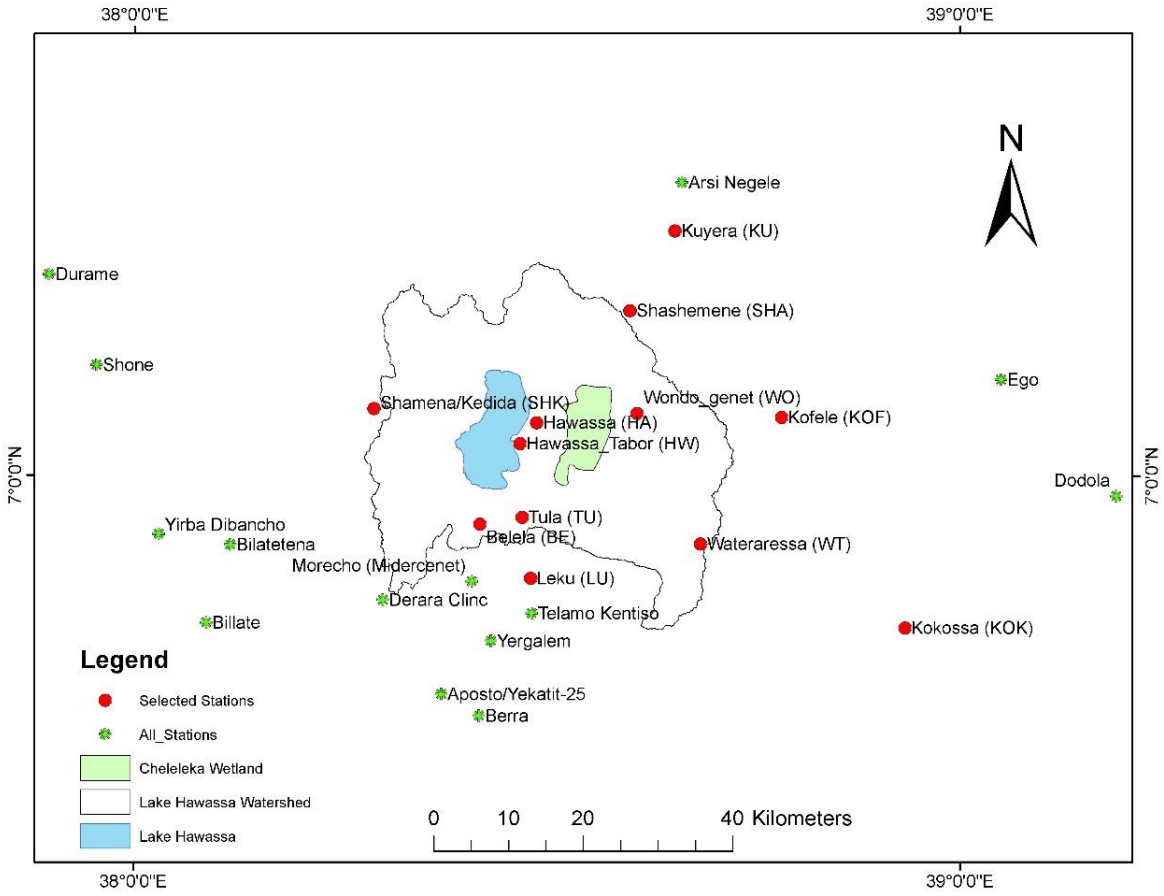


Figure 4 All rain gauge stations accessed for the study.

The evapotranspiration time series dataset is the other input required by WetSpa for simulation. After assessing the available meteorological records in and around the watershed, temperature-based evapotranspiration estimates were found to provide a better representation of the spatial variability in evapotranspiration. Hence, a simple temperature method (Eqn. 3) developed by Enku and Melesse (2014) was used to estimate evapotranspiration. However, due to the variability of the recording period, temperature data from only three stations (Hawassa, Wateraressa, and Kofele) were used.

$$ET_o = \frac{(T_{max})^n}{k} \quad \text{---------- Eqn. 3}$$

where ET_o is the reference evapotranspiration (mm day^{-1}), T_{max} is the maximum daily temperature (°C), n is a coefficient equal to 2.5, and k is a coefficient that can be estimated using $k = 48 * T_{mm} - 330$ for the combined dry and wet season, for the dry period $k = 73 * T_{mm} - 1015$ and for the wet season $k = 38 * T_{mm} - 63$. T_{mm} is the long-term daily mean maximum temperature for the seasons under consideration (°C).

Currently, satellite images are highly useful for estimating different meteorological parameters, such as potential evapotranspiration. As an input to assess the impact, GLEAM potential evapotranspiration estimates were also implemented in the third objective. GLEAM version 3.8a is a freely available ET estimate at daily, monthly and annual intervals with a $0.^{25^\circ}$ resolution. The daily GLEAM data for the years 2005-2014 were downloaded from the SFPT server (sftp://hydras.ugent.be), which can be accessed after providing the necessary information at http://www.gleam.eu/.

The GLEAM 3.8a dataset relies on reanalysis of radiation and air temperature, a combination of gauge-based, reanalysis, and satellite-based precipitation data, and satellite-based vegetation optical depth data (Martens et al., 2017). The components of evapotranspiration, such as transpiration (Et), interception loss (Ei), bare-soil evaporation (Eb), snow sublimation (Es) and open-water evaporation (Ew), are calculated separately, which later produce potential evaporation (Ep), root-zone soil moisture (SMroot), surface soil moisture (SMsurf), and evaporative stress (S)(Martens et al., 2017). In this study, one of the immediate outcomes, potential evapotranspiration was used.

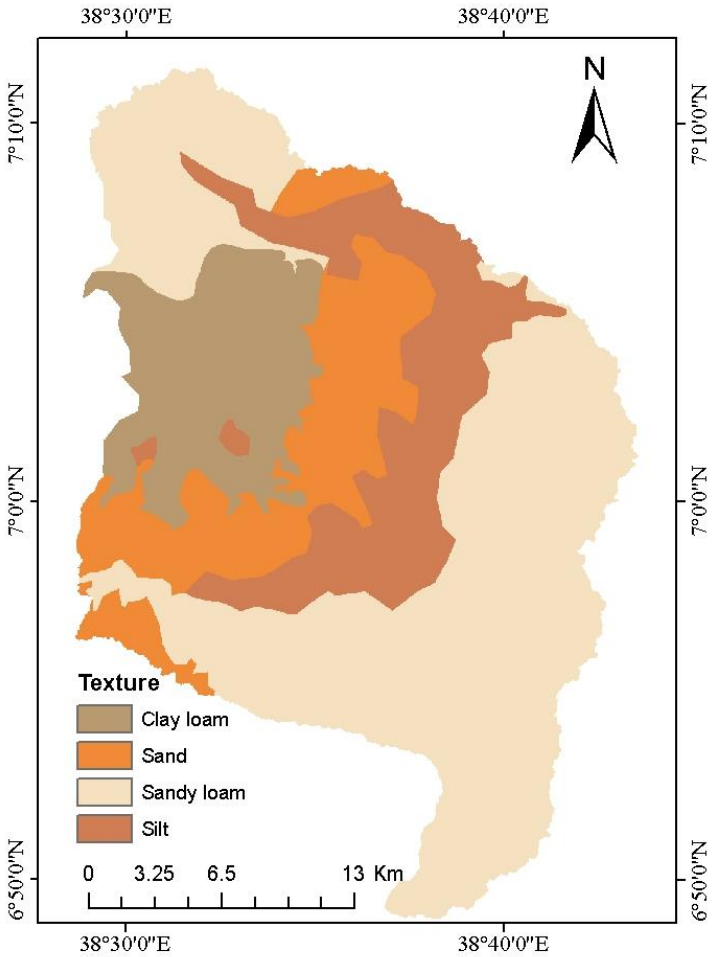


Figure 5 Soil texture map of the study area

Spatial datasets

The 30 m resolution DEM data were sourced from http://gdex.cr.usgs.gov/gdex/ and resampled to 32 m resolution for use as an input to the WetSpa model. Other spatial datasets input to the model include soil maps and land use/cover maps of the watershed. A soil map of the Tikur-Wuha watershed (Figure 5) was obtained from the Ethiopian Geological Survey Institute, whereas the land use/cover map was prepared by accessing the LandSat-8 image for the year 2014 and later classified by adopting the Google Earth Engine Code. The input raster pixel resolution and reference system of all spatial datasets were kept constant. Accordingly, a pixel resolution of 32 m was used, whereas a Universal Trans-Mercator (UTM) zone of 37 N was the constant reference system for each raster map.

Stream Flow Datasets

As explained in the previous section, the watershed consists of nine streams feeding the wetland. However, only two hydrological gauging stations are available at the subwatershed level, except for the two gauging stations available at the main river of Tikur-Wuha (i.e., downstream of the wetland). The first river gauge of Tikur-Wuha is located at the bridge near the inlet of Lake Hawassa, and another station, the Dato village station, is located upstream of the bridge. The stream record at the bridge location is available for the year before 2005 G.C., and during peak flood periods, the back water flow from the lake affects the gauge record. Currently, the bridge location is completely covered with vegetation, and the gauge is not functional. Hence, the discharge record at Dato village was used for calibration in this study.

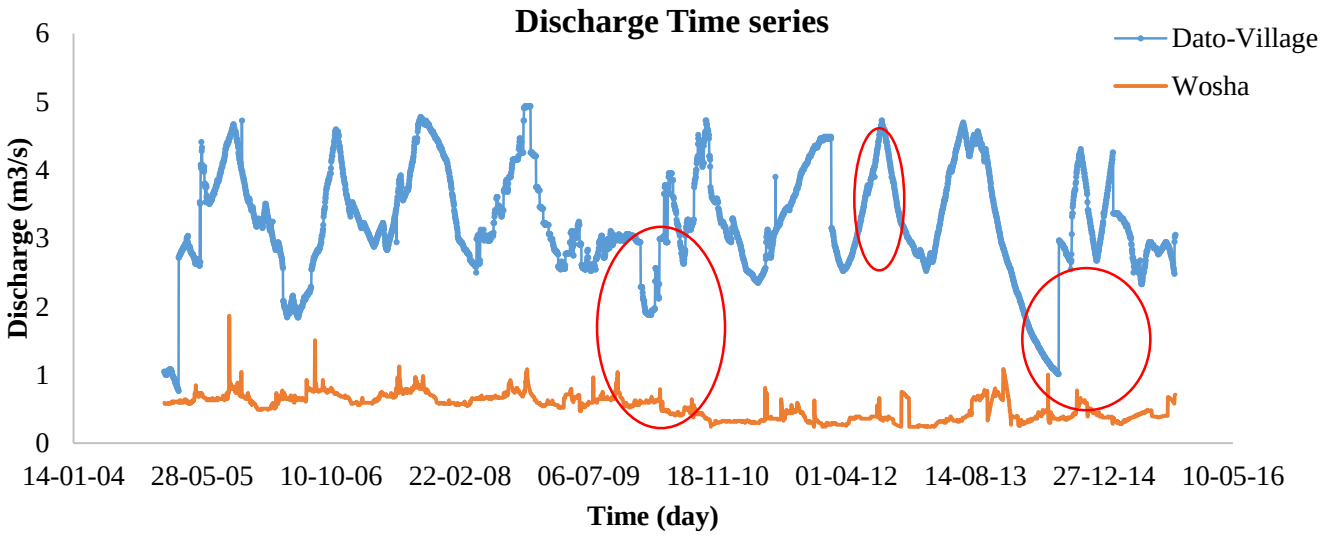


Figure 6 Discharge time series data

Other discharge records available at the subwatershed level are found at the Worka and Wosha stations. Discharge measurements at the Worka station are available only for the year before 2005 and are hence

disregarded in this study. Accordingly, the record of Wosha station (Figure 6) was used for calibration in addition to the record at the watershed outlet (i.e., Dato-Village).

Despite the fact that the WoSha River is a tributary of the Tikur-Wuha River, one may see a variation in the flow pattern by comparing the flow records. In-depth observation of the time series record (Figure 6) reveals that the peak flows and low flows are not similar to some portion of the record (i.e., red circles in Figure 6). Additionally, the sharp decrease in the recession limb and sharp increase in the rising limb of the hydrograph of the Tikur-Wuha River make it challenging to fully capture the flow pattern during modelling.

3.3. Methodology

3.3.1. Evaluation of a Multi-Staged Bias Correction Approach for the CHIRP and CHIRPS Rainfall Products

Ground measurements of rainfall are usually susceptible to errors, which places additional burdens on modelling efforts. Furthermore, ground observations are usually collected in accessible areas (i.e., close to urban areas and roads), which does not represent the distribution in remote and inaccessible areas (Dile et al., 2018, Dinku et al., 2014). This is where remote sensing data will be important. Unfortunately, these data also contain errors. Therefore, separate data are needed for bias correction and validation of satellite rainfall products (Gebremedhin et al., 2021). The ground observations are thoroughly assessed for any temporal or spatial inconsistency.

The satellite data were bias corrected and spatially disaggregated. Then, the available ground rainfall records are used for validation by dropping some stations at a time and reiterating each process. All the stations in the watershed, including those surrounding the watershed, are used. Bias correction was performed using a parametric empirical quantile mapping procedure. Downscaling was performed at a 1 km resolution by applying nearest neighbor (NN) and bilinear (BL) interpolation methods. Different methods have been introduced to merge satellite rainfall products with in situ observations. These include geographically weighted regression (GWR), conditional merging (CM), geographic differential analysis (GDA), and geographical ratio analysis (GRA).

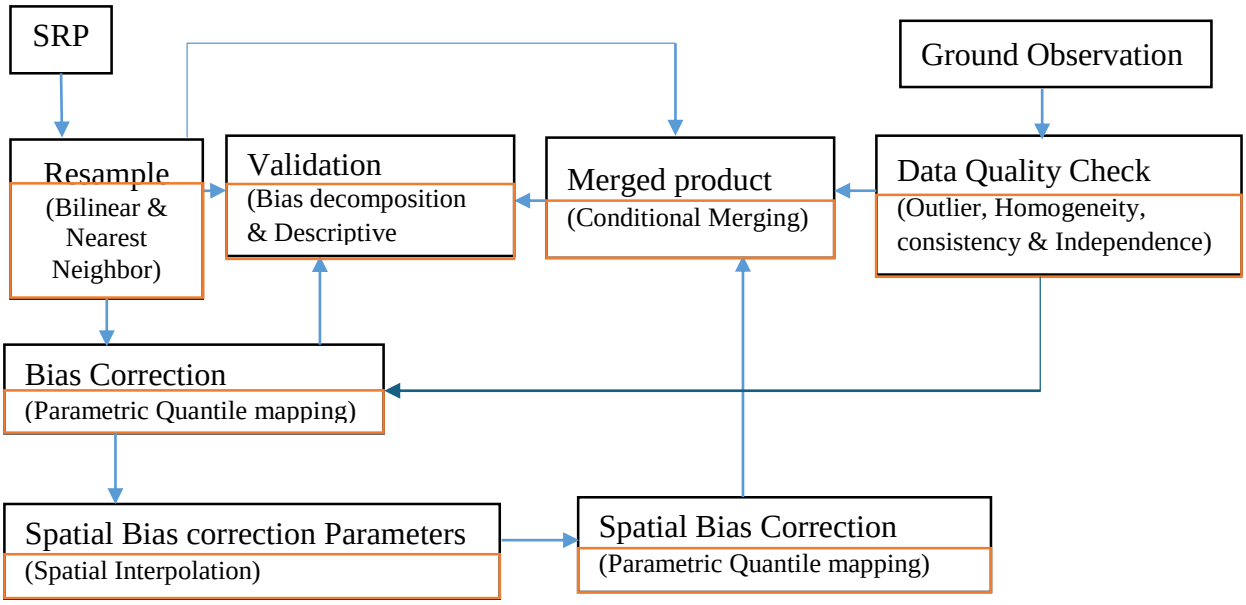


Figure 7 Flow chart of the general method; the methods implemented in each stage are shown in brackets

Based on the objective of the study, CM was implemented in this study. When applying the CM merging technique, both satellite data and in situ measurements at intersecting available grids were interpolated using the ordinary kriging method, and then the residual was added to the original satellite rainfall map to create the merged product (Jongjin et al., 2016). In general, this study applied a two-stage method (Figure 7); the first stage consisted of resampling the original SRP product to a 1 km resolution followed by bias correction, while the second stage used interpolated bias correction parameters at available grid locations to correct all grids and then merged with ground measurements via a conditional merging procedure.

One of the greatest challenges when working with large amounts of spatial and temporal data is the availability of comprehensive tools. There are multiple tools, such as ArcGIS and statistical software, to execute the individual and/or multiple processes mentioned above (Figure 7). However, it was challenging to find a tool that can execute all the tasks with minimum computation time and storage requirements. Hence, for this study, simple algorithms based on different existing modules written in the Python language were developed. Some of the open access libraries applied in these processes include the following: PYKRIGES, a kriging toolkit for interpolation tasks (Murphy, 2014); the Python Data Analysis (PANDAS) Library, a data analysis and manipulation tool including statistical analysis (McKinney, 2011); the Geospatial Data Abstraction Library (GDAL/OGR), a translator library for raster and vector geospatial data formats (GDAL/OGR Contributors, 2020); Numerical Python (NUMPY), for accessing, manipulating and operating on data in vectors and arrays (Harris et al., 2020); RASTERIO, for reading writing raster files (Gillies, 2019); XLRD, a library for reading data and formatting information from Excel files; SHAPELY,

for manipulation and analysis of planar geometric objects (Gillies, 2013); and Scientific Python (SCIPY), a library for optimization, mathematical calculations, statistics and others (Jones et al., 2001).

Bias correction and conditional merging algorithm

The developed algorithm is composed of four steps. The main stages of the program include preprocessing of the input data, station bias correction and fitting of the distribution parameter, spatial bias correction, and conditional merging. The main inputs to the program are raw daily satellite rainfall estimates in '.tif' file format and labeled with the date of the observation and daily gauge records prepared in '.csv' file format; furthermore, the length of the gauge record should be equal to the length of the SRP. Additionally, a simple graphical user interface (GUI) is prepared using the PyQt widget toolkit (Riverbankcomputing) and includes four separate tabs for each stage.

a) **Input data** preprocessing: At this stage, the program helps to prepare the satellite rainfall product. These include projection, creating boundary shape files, clipping, and resampling. Since most of these procedures require interpolation, the spatial reference system should be projected to metric systems, such as the transverse Mercator (UTM). However, clipping and boundary file creation are required to reduce the number of grid points to be processed.
b) **Bias correction and distribution fitting:** The pre-processed raster SRP files and ground observation data in the '.csv' file format are used as inputs for this stage, where a distribution is fitted to the monthly nonzero events at the collocated rain gauge locations. There are two widely used distribution options, Gamma and Exponential, where it is possible to choose both or either one. The gauge record is fitted to the selected distribution by excluding missing and nonzero events for all months separately. At this stage, the program is equipped to run bias correction (Eqn. 4 - 7), either including performance measures at each rain gauge location or without performance measures (Eqn. 11 - 15). The output at this stage includes fitted distribution parameters (i.e., rain gauge records) on a monthly basis, performance measure results, and bias-corrected rainfall data at all gauge locations. (Refer to Annex 1)
c) **All Grid Bias Correction:** Using the boundary extent, fitted distribution parameters are interpolated by applying the ordinary kriging technique for all months separately. The mapped/gridded parameters are then used to conduct bias correction at all grid points for the projected raster SREs. A K-fold validation process is also optional in this stage. Hence, the working extent, resolution, number of validation stations at each iteration, and projection method need to be specified in this stage. Otherwise, one can also choose to only perform spatial bias correction

without the performance measures. The outputs include the corrected raster file and the performance indicators in the '.csv' file of each iteration. (Refer to Annex 2)

d) **Conditional Merging:** The corrected raster file and rain gauge records are used as inputs in this stage. With the specified procedure (i.e., section below), conditional merging is conducted using the inputs. As a cross-validation mechanism, the program allows one to randomly drop a specified number of stations and conduct the merging process while performing validation for the hidden stations. The process can be repeated for a specified number of iterations. For practical application, the program can also be conducted without a performance check. Based on the choice of simulation, the output includes merged raster files and performance measures for each iteration. (Refer to Annex 3)

Resampling

Recent studies have recommended the use of information gathered from multiple sensors and/or satellites to increase the likelihood of retrieving the best representation from different sources (Dinku et al., 2011, Dubovik et al., 2021). However, incompatible scales and resolutions of different satellite observations are major challenges in most retrieval algorithms (Omondi, 2017). With respect to the methods applied, resampled raster data usually suffer from the inability to maintain the originally stored information (Omondi, 2017). Hence, one should expect that different methods might have different error propagation patterns.

One of the widely utilized resampling methods is the NN resampling method. This method is advantageous because of its lower computational time requirement, ease of application, and ability to maintain the original stored information (Brandsma and Können, 2006). However, Omondi (2017) indicated that this method is the least accurate method because it distorts results and sometimes omits or duplicates multiple cell values. Complemented for its ability to smoothen the output raster and is usually recommended for upsampling, BL resampling methods are based on the core method of averaging values in the nearest four grids (Baboo and Devi, 2010).

Rain gauge observations are usually perceived as the true value of rainfall and are used for correcting satellite rainfall products. In contrast, one can argue that most ground observations are full of errors, especially when manual rain gauges are used in developing countries (Beyene et al., 2018, Dile et al., 2018). Aside from these modest assumptions, spatial mismatch (i.e., between rain gauge point observations and pixel values of satellite rainfall products) is another eminent source of error, especially for low spatial resolution (large pixel size) products (Dinku et al., 2011, Gebere et al., 2015). As a solution to this challenge, many studies (Gebremedhin et al., 2021, Jongjin et al., 2016) have depended on interpolating ground

observations to create matching gridded rainfall maps. Readers are forwarded to reviews Hu et al. (2019) and Li and Heap (2008), where for an elaborate discussion on available methods of interpolation, including the pros and cons of the methods. Conducting the merging process requires creating similar resolutions between the SRP and interpolated ground station measurements. Hence, both the NN and BL resampling methods were tested for performance and use in this study.

Bias Correction

Bias correction based on the sampling window and length of the sampling window are well-practiced techniques for improving bias correction (Bhatti et al., 2016). To decrease the percentage of wet spills in the CHIRP SRE, this study considered 3- and 1-day central moving windows for dry and wet periods, respectively. After extracting the estimated rainfall at a specified grid point (j) via the pixel-to-point approach, the algorithm calculates the air distance from the center of the grid to all available rain gauge station locations. Thereafter, assuming rain gauge density requirements of 1 in 100 km^2, if the minimum distance to any of the stations falls below 10 km, that station (x) is selected for comparison (Eqn. 4). On the other hand, if the distance exceeds 10 km, the four closest stations are selected for comparison.

$$\begin{Bmatrix} \boldsymbol{P_{sj}}^{\boldsymbol{i}} = \boldsymbol{0} & \boldsymbol{if} & \boldsymbol{P_{gx}}^{\boldsymbol{i}} = \boldsymbol{0} \\ \boldsymbol{P_{sj}}^{\boldsymbol{i}} = \boldsymbol{P_{sj}}^{\boldsymbol{i}} & \boldsymbol{if} & \boldsymbol{P_{gx}}^{\boldsymbol{i}} > \boldsymbol{0} \end{Bmatrix} \text{ ---------- } \textit{Eqn. 4}$$

where ${P_{gx}}^{i} = \left[{P_{g1}}^{i}, {P_{g2}}^{i}, \ldots, {P_{gn}}^{i}\right]$, for a given day (i), if any of the rain gauge stations (x) record rainfall (${P_{gx}}^{i}$), the estimated SRE (${P_{sj}}^{i}$) will be maintained; if no rainfall is detected by any of the selected stations, the estimated SRE is replaced by zero. The resulting rainfall time series are used for bias correction. However, this procedure was not performed on the CHIRPS product, realizing that the estimates are already corrected using few ground rain gauge records.

The distribution mapping (DM) technique was originally developed to bias correct regional and global climate models (McGinnis et al., 2015, Switanek et al., 2017). One of these DM techniques, quantile mapping (QM), has been proven to correct both regional and global climate circulation modes (Heo et al., 2019, Teutschbein and Seibert, 2012). Additionally, QM techniques are chosen for bias correction of satellite rainfall products because of their ability to adjust the standard deviation of daily satellite rainfall products while preserving other moments (Katiraie-Boroujerdy et al., 2020).

QM is based on parametric empirical distribution mapping (Eqn. 5, 6, & 7), and the method of bias correction was applied in this study. As presented by Themeßl et al. (2012), the corrected rainfall $(R_{t,i}^{cor})$ at grid (i) and day (t) is given as follows:

$$\boldsymbol{R_{t,i}^{cor} = R_{t,i}^{raw} + Cf_{t,i}} \text{ ----------Eqn. 5}$$

$$\boldsymbol{Cf_{t,i} = ecdf_{m,i}^{obs,cal^{-1}}(P_{t,i}) - ecdf_{m,i}^{mod,cal^{-1}}(P_{t,i})} \text{ ---------- Eqn. 6}$$

$$\boldsymbol{P_{t,i} = ecdf_{m,i}^{mod,cal}(R_{t,i}^{raw})} \text{ ---------- Eqn. 7}$$

where $Cf_{t,i}$ is the transfer function; $ecdf_{m,i}^{obs,cal^{-1}}$ and $ecdf_{m,i}^{mod,cal^{-1}}$ are the observed (obs) and modelled (mod) inverse empirical months (m), respectively; and P is the probability during the calibration (cal) period.

For a given station and month, the ground observation (without missing data) and resampled SRP at the coinciding grid point were fitted to an exponential distribution using the maximum likelihood method. Additionally, the distribution was fitted to nonzero events, and missing data were excluded. Furthermore, a standard check of the chi-square test and Kolmogorov–Smirnov test was conducted to confirm a good fit of the distribution to each dataset. Once the distribution parameters at the collocated grid points were determined, the parameters were transferred to grid points with no ground observations through interpolation (ordinary kriging). Hence, bias correction was conducted for all grid points with and without ground control observations. This would not have been possible with the nonparametric QM technique. Additionally, parametric quantile mapping methods allow the correction of maximum events (not recorded by ground stations) that may require extrapolation while transferring distribution parameters.

Conditional merging

Implementing a merging technique using two sources enables the capture of the complementary merits of each source, which could reduce bias in satellite rainfall products (Jongjin et al., 2016). Conventional merging, which is based on ordinary kriging (OK) interpolation, is a merging technique that is used to merge radar data and meteorological data (Guenzi et al., 2017). Ordinary kriging is a regression technique that allows for value interpolation while minimizing the mean squared error (Guenzi et al., 2017). While performing CM, it is anticipated that a theoretical semivariogram based on datasets of regularly distributed precipitation will be fitted to the experimental semivariogram for the operation of OK (Jongjin et al., 2016).

The experimental semivariogram was calculated and then fitted to the theoretical semivariogram using weighted least squares to determine the appropriate variogram parameters, such as the nugget, sill, and range (Jongjin et al., 2016). In this study, CM was applied following the procedure applied by (Jongjin et al., 2016). The initial step was the interpolation of the ground station record to a 1 km resolution field using OK. The second step required the extraction of the collocated (i.e., with ground observation stations) resampled (1 km resolution) SRP grid cell observations. The third step was the interpolation of the extracted

SRP grid cells to a 1 km resolution field using OK. The fourth step error was calculated by deducting the rainfall map created in the third step from the resampled SRP observation. Finally, this error field was added to the rainfall map created in the first step.

The merging procedure was conducted in two stages. The first stage was before applying the bias correction technique to the original SRP dataset. Next, it was applied after applying the bias correction described in the previous stage. Both products were tested for performance, in which the necessity of the procedure devised in this study is significant.

3.3.2. Uncertainty assessment method

The uncertainty assessment in this study was conducted by adopting three step simulations (Xie et al., 2018)on predefined model setups. The model setups were created by only varying the input rainfall and potential evapotranspiration data. The first model setup was for input rainfall from ground observation. The second model setup used bias-corrected and conditionally merged CHIRP satellite rainfall product. The third model setup was based on the GLEAM ET estimates.

The next step were defining upper and lower limits for global parameters of WetSpa model and the first uncertainty assessment procedure, which is creating 10,000 possible parameter combinations were conducted. A Monte Carlo simulation stratified by Latin Hypercube sampling technique was used to create the parameter combinations. A total of 30,000 semi-distributed simulation runs of WetSpa was executed for the three model setups. By keeping all other inputs or set up of WetSpa model the assessment of recharge uncertainty due to input rainfall and/or evapotranspiration was individually investigated. Hence, the only simulation difference between the three model setups was the input rainfall and evapotranspiration data.

The second step was to define and select acceptable model realisation for each model setup run. This was achieved by selecting the 100 best simulations with the least root mean squared error values and the percent of volume error between the observed and simulated discharge. For the third model setup additional cut-off value of $0.9 m^3/s$ for RMSE was used to select the acceptable model realisations. Hence, the study used arbitrary cut-off values, ensuring the consistency of the simulation and ensuring sufficient observations for further statistical analysis. The selected 100 parameter combinations were used to simulate the fully distributed groundwater recharge of the watershed. The final step of the uncertainty assessment was to define uncertainty using descriptive statistics such as the standard deviation of the long-term mean annual recharge resulting from the accepted model realizations.

3.3.3. Assessment of Input Rainfall Uncertainty for Groundwater Recharge Estimation

Uncertainties from the input rainfall data were quantified first by creating two separate model setups. This was achieved by keeping all inputs to the selected hydrological models similar and by changing the input rainfall time series. The input time series to the first model setup (M_1) was prepared from available rain-gauged records, which were subsequently converted to areal rainfall by using the Thiessen polygon. In the second model setup (M_2), rainfall data were prepared by first bias correcting (i.e., using empirical parametric quantile mapping) the CHIRP SRE and later merged with available rain gauge data via the conditional merging technique.

The next step of the method requires bounding the global parameters of the selected hydrological model based on prior watershed knowledge and a literature review. Then, Monte Carlo simulation stratified by the Latin hypercube sampling approach was conducted to generate 10,000 possible parameter (i.e., WetSpa model parameter) combinations using the R programming language. Second, semi-distributed model simulations using all parameter combinations were applied for each model setup separately, and the model performance (root mean squared error (RMSE) and percent error in volume (PVE)) was determined. The third step was quantifying the long-term mean recharge for each accepted model realization, and the statistical long-term mean recharge distribution was analysed.

Due to the scarcity of overlapping hydrometeorological data, for this study, only daily time step simulations were conducted for 10 consecutive years (2005-2014). Additionally, from the accepted model realizations, 100 best parameter combinations were selected, and distributed model simulations were conducted separately for each model setup. The resulting 100 recharge maps per model setup were considered to produce spatial statistics and related uncertainty in recharge estimation. The number of simulations was limited due to the computational capacity requirement. Furthermore, Xie et al. (2018) showed that 10,000 simulation runs per model setup is sufficient to capture acceptable variability in parameter combinations. Each model setup simulation run (i.e., semi-distributed) was automated using the Python programming language, whereas the fully distributed model simulation was conducted manually.

The semi-distributed model simulation produces water balance components at a daily time step, whereas the fully distributed model simulation produces one gridded recharge map over the simulation time period (i.e., 10 years for this study). Hence, it was necessary to determine the long-term mean annual recharge for each semi-distributed model simulation by first aggregating the daily recharge for each year and then averaging it annually, whereas for the fully distributed simulation, each cell was divided by the length of the simulation period to acquire the mean annual recharge for each model simulation.

Land Use/Cover Classification

A supervised land use classification was used to create land use map of the study area. The classification was done by using LandSat-8 image for the year 2014. Freely available Google Earth engine code at https://developers.google.com/earth-engine/guides/classification for supervised land use classification was adopted. The supervised classification procedure involves screening images for cloud cover, correcting for surface reflectance, collecting training points (i.e., ground truth points of each land use type) and classification. The training points (i.e. Crop land (110 points), Water body (32 points), Grassland (80 points), Permanent wetland (22 points), Mixed forest (52 points), Urban and built-up area (65 points), Barren or sparsely vegetated land (7 points), & Open shrub land (43 points)) for the classification were collected using Google earth map. The selected training points are randomly split into two sets for calibration and validation, where the smile random forest classifier (Hasan et al., 2022) is applied for classification. The adopted code allows for the computation of training and validation accuracy. A confusion matrix representing substitution accuracy was constructed, and finally, the kappa coefficient (Kraemer, 2014) was calculated to assess the accuracy of the classification.

WetSpa Model

The hydrological simulation in this study was conducted using a physically based and distributed hydrological model for predicting the water and energy transfer between soil, plants and atmosphere (WetSpa) at daily time steps. Specifically, the WetSpa extension (Liu and De Smedt, 2004), which is a GIS-based distributed watershed model, was employed in this study. WetSpa simulates hydrological processes (precipitation, interception, snowmelt, depression, infiltration, evapotranspiration, percolation, surface runoff, interflow and groundwater flow) at the grid cell level by dividing the hydrological system into four control volumes (i.e., plant canopy, soil surface, root zone, and groundwater aquifer) (Liu and De Smedt, 2004).

The hydrological simulation using the WetSpa extension is generally achieved in three stages. The first step requires the determination of local variables based on the input gridded DEM, soil map, and land use map. This step was conducted using the ArcView 3.2 program. All local variables are predetermined based on the three base maps. The readers are referred to Liu and De Smedt (2004) for a detailed description of the preparation process and model run using the WetSpa extension program. The second step is to run the external executable, where the extension provides a choice of fully distributed and semi-distributed model run types.

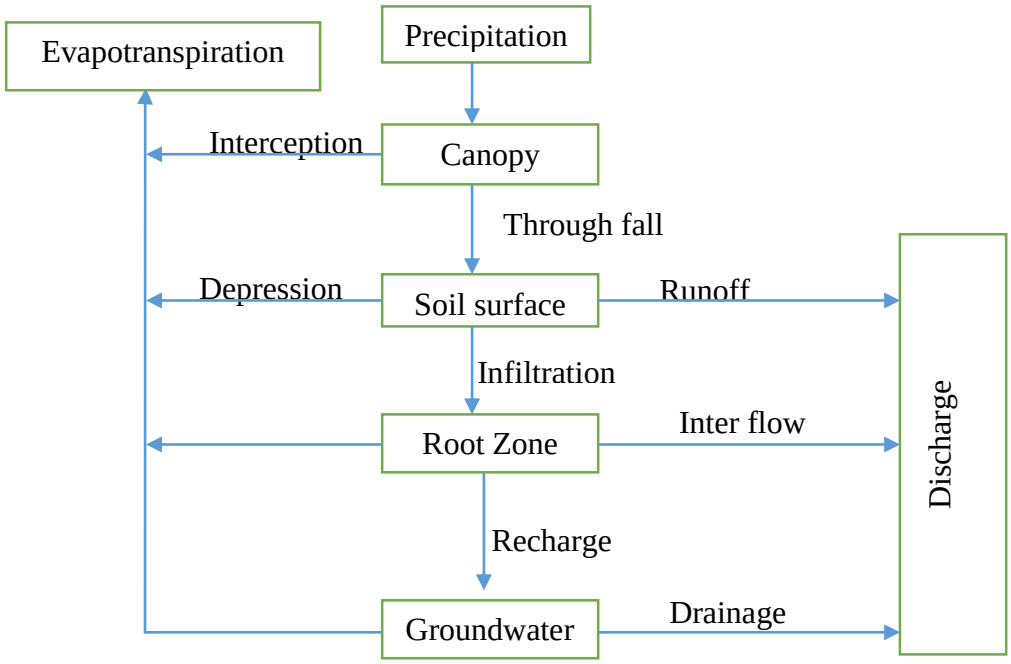


Figure 8 WetSpa simulation at the cell level (Liu and De Smedt, 2004)

The model requires input point rainfall records from available rain gauges in and around the watershed of interest. These point rainfall records are converted to areal rainfall estimates using the Thiessen polygon method. For the fully distributed model run, the developed Thiessen polygon map is converted to a raster map, where the station number is the cell value. For this study, a similar procedure was followed to create the Thiessen polygon map using the available rain gauge stations for the M_1 model. For the M_2 model setup, each grid cell center was assumed to represent the rain gauge station, and the grid map was considered the Thiessen polygon.

WetSpa parameter range

WetSpa simulations (Figure 8) are conducted at the cell level; hence, all local parameters are determined for each cell. The 32 × 32 resolution resulted in 859 columns with a 1296-row grid mesh area for the model. Based on the input soil texture class, land use type, and physiographic characteristics of each cell, the default local parameters are assigned by the WetSpa extension program. All default parameters were kept except for the change in root depth for forests from 1 m to 5 m based on the dominance of eucalyptus trees in the watershed forest cover. Similar changes were made by the Yenehun et al. (2022) application of the WetSpass in the Lake Tana basin and the Yenehun et al. (2020) in the Gilgel Abay watershed.

Table 4 WetSpa global parameter range

Parameter	Range	Description
Ki	1.5 – 2.2	Seasonally vegetated farm lands and commercial mixed forest are the dominant land covers
kg	0.0001– 0.0003	Based on the recession curve of observed hydrograph
K-s	1.0 – 1.2	Assuming up to 20% influence of initial soil moisture condition
K-ep	0.65 – 0.95	Based on the assumption that can represent seasonal variability
g0	1000 – 1900	Based on the observed annual rainfall amount
g-max	1000 – 5100	Based on the assumed maximum root depth
k-run	7.5 – 10.5	The observed hydrograph has low noises
P-max	100 – 200	Based on the maximum observed rainfall and error factors

There are 11 global parameters (Table 4) used for the calibration of model runoff production and hydrographs at watershed outlets (Liu and De Smedt, 2004). Three of these parameters (base temperature for snowmelt (To), temperature degree-day coefficient (k-snow), and rainfall degree-day coefficient (k-rain)) are assigned to control the processes of snow melt (Bahremand et al., 2021). For regions with no snow occurrence, these parameters can be set to -1.0 (Liu and De Smedt, 2004). Similar adjustments were made by Kidanemariam et al. (2021) by considering the formation of precipitation in the focus basin. Adjusting less sensitive or insensitive parameters is an added advantage in narrowing the feasible parameter range (Bahremand, 2016). Another less sensitive parameter is the initial soil moisture condition (Ks), which only affects the initialization period of the simulation (Bahremand et al., 2021).

Other parameters can be assigned based on the physical and meteorological characteristics of the watershed under study (Bahremand, 2016). Accordingly, the base flow recession curve coefficient (kg) can be determined from the slope of the master curve of the recession limbs; the observed precipitation can be used to determine the feasible range of the correction factor for potential evapotranspiration (K-ep) and the surface runoff exponent for a near zero rainfall intensity (K-run). Additionally, the cumulative flow curve can indicate the feasible maximum groundwater level (g-max), while annual rainfall can be a good indicator of initial ground water storage (go)(Bahremand et al., 2021). Finally, the scaling factor for interflow (Ki) can be bounded by looking at the dominant land use type of the watershed (Bahremand et al., 2021).

3.3.4. Evaluation of GLEAM potential evapotranspiration data and related groundwater recharge estimation uncertainty

Research (Xie et al., 2018, Beyene et al., 2024) conducted to analyse the impact of input evapotranspiration data on the uncertainty of groundwater recharge estimates has proven that the use of good-quality ET estimates can reduce the uncertainty of recharge estimates. A poor distribution of meteorological measurements in watersheds such as Tikur Wuha and a high rate of missing data can create significant uncertainty in the estimation of potential evapotranspiration. For instance, a missing data rate as high as 36% at the Kokosa meteorological station (Table 3) was detected for the recorded precipitation. It is no different for the other meteorological stations. Additionally, the sparsely populated first-class meteorological stations in and around the Tikur Wuha watershed impact the quality of potential evapotranspiration estimates. Hence, an additional freely available satellite and reanalysis-based potential evapotranspiration estimate called GLEAM was evaluated, where it is hypothesized that such data inputs can reduce the uncertainty of groundwater recharge estimates.

To investigate the impact of multiple data sources of input evapotranspiration on the uncertainty of groundwater recharge estimates, the general methodologies and input data discussed in section 3.3.2 were kept unchanged. These datasets include all the input datasets, such as the DEM, land use/cover map, and soil map. Additionally, the simulation was conducted by using the WetSpa model, where the previously defined parameter ranges are also kept constant for this simulation. However, the simulation was only conducted by using the model set up prepared for satellite-based rainfall estimates. Hence, a third model setup (M3) was prepared by changing only the input ground-based potential evapotranspiration estimates with the satellite-based estimates of the GLEAM dataset.

The dataset of GLEAM potential evapotranspiration data was downloaded for the years 2005-2014 on a daily scale and at a special scale of $0.^{25^{\circ}}$. Converting the NETCDF file and pre-processing of the data were conducted by using packages in the R program. After extracting the dataset, a Thiessen polygon resembling the original grids of the GLEAM dataset was developed as an input to M3. Finally, a separate 10,000 semi-distributed model simulation was conducted using the previously randomly generated WetSpa parameter combinations. The uncertainty of the model was analysed by evaluating the distribution of the RMSE results for the 10,000 simulations.

3.4. Performance measure

The validation was conducted at several stages of bias correction, which enabled us to determine the importance of each process. The objective function measures the goodness of fit between the computed and

observed data for a selected element. The choice of the objective function depends upon the need. Both bias decomposition and statistical methods were applied for two sets of data periods (i.e., wet period and dry period). The following statistical measures were used to quantify the accuracy of bias correction outputs. These include the percent error in volume (PVE), coefficient of determination (r^2), and root mean squared error (RMSE), which are widely applicable in hydrologic modelling.

The coefficient of determination (r^2) describes (Eqn. 8) the degree of collinearity between bias-corrected and gauged data (Moriasi et al., 2007).

$$\boldsymbol{r^2} = \frac{\left(\sum(\boldsymbol{P_{gi}}-\overline{\boldsymbol{P_g}})\ \sum(\boldsymbol{P_{si}}-\overline{\boldsymbol{P_s}})\right)^2}{\sum(\boldsymbol{P_{gi}}-\overline{\boldsymbol{P_g}})^2 \sum(\boldsymbol{P_{si}}-\overline{\boldsymbol{P_s}})^2} \text{ ---------- } \textit{Eqn. 8}$$

where r^2 is the coefficient of determination (where values can range from -1 to +1), P_{gi} is the gauged rainfall value at the i^{th} time interval, P_{si} is the SRE at the collocated grid at the i^{th} time interval, $\overline{P_g}$ is the mean gauged rainfall and $\overline{P_s}$is the mean SRE at the collocated grid. A perfect positive correlation of rainfall estimates is indicated by a value of 1, while a value of 0 implies no correlation.

The PVE ranges from –ve ∞ to +ve ∞, where results within -10% to +10% indicate good agreement and results out of the indicted range show poor performance, whereas 0 is the perfect match. The PVE (Eqn. 9) considers the computed volume and does not account for the magnitude of the peak;

$$PVE = 100\left|\frac{Vo-Vs}{Vo}\right| \text{ ---------- } \textit{Eqn. 9}$$

where PVE is the percentage error in volume, Vo is the observed volume and Vs is the simulated volume.

The error variation between the raw and bias-corrected products against ground observations can be indicated by the RMSE (i.e., ranging from 0 to ∞), where close to zero results indicate very good performance and vice versa. The mathematical representation of the root mean squared error (RMSE) (Eqn. 10).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(P_{si} - P_{gi}\right)^2} \text{ ---------- } \textit{Eqn. 10}$$

where P_{si} & P_{gi} are the satellite rainfall product and ground-based rainfall observations on day i, respectively, and n is the total number of available pairs of data. A performance check of each output was conducted by using descriptive statistics and bias decomposition. The classification of error components in satellite rainfall products includes hit bias (HB), miss bias (MB), and false alarm/precipitation (FA) methods. HB is the sum of differences for the over/underestimation of ground observations detected by

satellites, whereas MB is the sum of all undetected gauge events detected by satellites, while FA is the number of all events estimated by satellites but not by gauging stations (Omondi, 2017).

For a better description of the results from satellite observations compared to ground observations, bias decomposition methods such as hit bias (HB), miss bias (MB), percentage of detection (POD), false alarm ratio (FAR), and frequency bias index (FBI) were applied.

$$HB = \sum_{i=1}^{n}(P_{si} - P_{gi}) : (P_{si} > 0 \ \& \ P_{gi} > 0) \text{-----} \textit{Eqn. 11}$$

$$MB = -\sum_{i=1}^{n} P_{gi} : (P_{si} = 0 \ \& \ P_{gi} > 0) \text{ ----------} \textit{Eqn. 12}$$

MB and HB measure the total depth of missed and correctly detected rainfall events, respectively. For these measures, results close to zero indicate very good agreement, and the result (i.e., HB) can range from –ve ∞ to +ve ∞, whereas MB values range from –ve ∞ to 0.

$$POD = \frac{H}{H+M} \text{ ----------} \textit{Eqn. 13}$$

$$FAR = \frac{FA}{H+FA} \text{ ----------} \textit{Eqn. 14}$$

$$FBI = \frac{H+FA}{H+M} \text{ ----------} \textit{Eqn. 15}$$

These indicators depend on the number of SREs, successfully detected events (H), missing SRE events that are measured by rain gauges (M), and rainfall detected by the SRE but not measured by a rain gauge (FA). The percent of detection (POD) calculates the ratio of observed rainfall rates that are correctly detected by satellite rainfall estimates (SREs) (H), whereas the false alarm ratio (FAR) calculates the ratio of wrongly detected rainfall rates (FA) by SREs. The ideal scores for the POD and FAR indices, which range from 0 to 1, are 1 and 0, respectively. The frequency of bias index (FBI) calculates the ratio of the total number of rainfall events observed to the total number of rainfall events detected by the SREs, with a value ranging from 0 to ∞. FBI values greater than 1 suggest overestimation by SREs, whereas FBI values less than 1 imply underestimation, with 1 being the perfect value.

Calibration procedures can help narrow the appropriate parameter range for a given watershed. Hence, the Nash Suit cliff efficiency (NSE) (Eqn. 16) was implemented in the calibration process, where a value of 1 indicates a perfectly fit model.

$$\boldsymbol{NSE = 1 - \left(\frac{\sum_{i=1}^{n}(sim-obs)^2}{\sum_{i=1}^{n}(obs-\overline{obs})^2}\right)} \text{ ---------- } \textit{Eqn. 16}$$

where NSE is the Nash and Sutcliffe efficiency; obs is the observed value at the i^{th} time interval; sim is the simulated value at the i^{th} time interval, and $\overline{obs}$ is the mean of the observed parameter.

One additional performance measure used for evaluating the outcomes of WetSpa model simulation results is the model volumetric efficiency (MVE). The value of MVE (Eqn. 17) ranges from 0 to 1, where 1 indicates a perfect match.

$$\mathrm{MVE} = 1 - \left(\frac{\sum_{i=1}^{n}|Q_{si} - Q_{oi}|}{\sum_{i=1}^{n} Q_{oi}}\right) \text{ ---------- } \textit{Eqn. 17}$$

where the observed discharge at time step i is Q_{oi} and Q_{si} is the simulated discharge at time step (i).

CHAPTER FOUR

4.0. Results and discussion

4.1. Evaluation of a Multi-Staged Bias Correction Approach CHIRP and CHIRPS Rainfall Products

The raw CHIRP and CHIRPs products were analysed, and the results showed that the older/prior version had a greater percentage of wet spills or erroneous detections. As a result, the detection percentage of the CHIRP version is higher. The CHIRP rainfall estimates have a greater than 50% false alarm ratio.

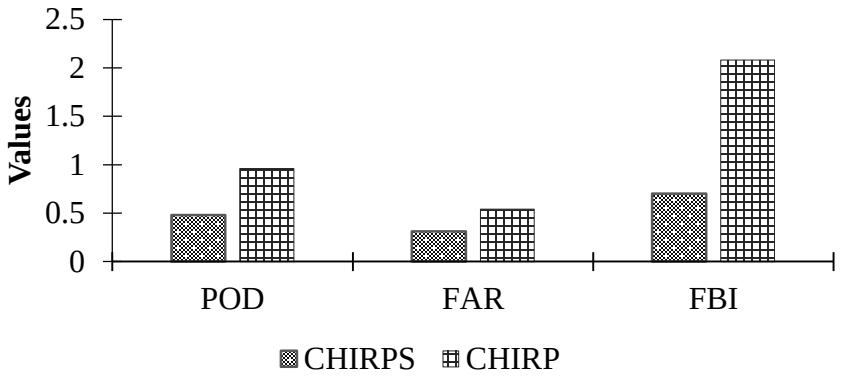


Figure 9 Comparison of the raw CHIRP and CHIRPS datasets at the Hawassa gauging station

As an illustration of the results obtained (Figure 9) at the Hawassa rain gauge station (i.e., using point-to-pixel comparison), 95% of the rainfall recorded by the ground station was detected by the CHIRP, whereas more than 50% of the wet spills were false alarms. However, for the false bias index values, the deviations from the ground station are 31% and 100% for CHIRPS and CHIRP, respectively. However, due to the inability of the index to assess correctly detected SREs that are not measured by ground stations, one cannot draw any firm conclusions about performance based only on the FAR.

4.1.1. Resampling

The results from the applied spatial disaggregation methods (bilinear and nearest neighbor) were compared to the corresponding available grid points with ground observations. Examining the NN resampling/downscaling method for the dry period (Table 5), where the raw SREs are preserved, the percentage of CHIRPS (V2) detection falls below 30% for most locations. An FAR as high as 69% (Annex 4) indicates that the performance is poor. Similar conclusions were made by Goshime (2020), who reported more than 50% false detections. However, the descriptive statistical results show that very good agreement was shown at most stations, especially at the Leku, Hawassa, and Hawassa_Tabor stations.

*Table 5 Performance evaluation of NN resampling (*i.e.*, to 1 km resolution) techniques for dry and wet periods for CHIRP (V0) and CHIRPS (V2).*

Station	Wet period				Dry period			
	PVE (%)		POD		PEV (%)		POD	
	Vo	V2	Vo	V2	Vo	V2	Vo	V2
BE	-21.1	-19.7	0.98	0.45	-89.1	-60.3	0.77	0.29
HA	-3.38	1.81	0.98	0.41	-16.8	9.49	0.86	0.3
HT	-15.3	-9.8	0.98	0.44	-28.5	-5.9	0.85	0.31
KOF	-7.8	-5.3	0.99	0.38	-0.26	14.3	0.9	0.26
KOK	35.6	38.5	0.99	0.33	50.7	57.3	0.83	0.21
KU	-21	-19.3	0.93	0.36	-56.2	-18.7	0.89	0.28
LE	-25.9	-24.2	0.98	0.46	-13.2	3.02	0.83	0.35
SHK	-42.7	-37.1	0.98	0.4	-50.9	**-17.9**	0.81	0.29
SHA	-97.2	-87.2	0.98	0.37	-95	**-46.2**	0.83	0.25
TU	-20.3	-27.7	0.97	0.44	3.12	11.2	0.82	0.29
WT	-16.1	-13.4	0.97	0.38	-39.9	-21.2	0.86	0.23
WO	1	4.9	0.98	0.41	-4.23	17.7	0.86	0.26

On the other hand, the performance of CHIRP (V0) was very good, according to the RMSE (Annex 4) and POD (Table 5) results. Nevertheless, as illustrated in the above section, this is due to the higher rate of wet spill in the V0 SREs than in the V2 SREs. Very high FAR results are recorded in V0 in comparison to V2 because V2 is sourced from ground observations and the SRP (Funk et al., 2015). A negative percentage error in volume indicates overestimation, while nearly zero results imply very good agreement. The overestimation in volume reached as high as 89% (at the Belela station) using the NN resampling technique, whereas the BL method reduced the overestimation by 2% (Table 6).

*Table 6 Performance evaluation of wet period BL resampling (*i.e.*, to 1 km resolution) techniques for CHIRP (V0) and CHIRPS (V2).*

Station	Wet Period				Dry Period			
	PEV (%)		POD		PEV (%)		POD	
	Vo	V2	Vo	V2	Vo	V2	Vo	V2
BE	-18.5	-17.4	0.98	0.51	-87.3	-58.8	0.77	0.31
HA	-5.94	-0.19	0.98	0.51	-19.1	6.54	0.86	0.4
HT	-12.7	-8.38	0.98	0.51	-25.4	-4.2	0.84	0.37
KOF	-8.01	-5.29	0.99	0.46	0.49	15.01	0.9	0.35
KOK	35.6	37.48	0.99	0.41	50.76	57.3	0.83	0.26
KU	-20.4	-17	0.93	0.46	-56.1	-10.9	0.89	0.33
LE	-25.2	-23.8	0.98	0.51	-13.4	2.88	0.83	0.38
SHK	-41.6	-35.9	0.98	0.47	-50.6	-18.8	0.81	0.38
SHA	-95.1	-91.7	0.99	0.44	-95.7	-54.5	0.88	0.31
TU	-24	-23.7	0.97	0.49	0.54	14.08	0.82	0.34
WT	-16.8	-13.8	0.97	0.47	-40	-21.5	0.9	0.3
WO	1.56	5.1	0.99	0.5	-2.72	18.54	0.86	0.31

A slight improvement in the performance (i.e., descriptive statistics) of the two resampling methods was shown by the BL SREs. However, examining the bias decomposition result, the NN showed slight performance improvement. These findings are consistent with the core principle of resampling techniques, where new and undetected values are assigned by the BL technique from neighboring grids (Omondi, 2017). However, the NN method preserves/does not introduce new observations at focus grids. For the study period (i.e., 11 years), the total depth of the missing and correctly estimated SRP at the collocated ground station is determined by the MB and HB (Annex 6 & 7). A higher positive HB value (during the wet period at the Shashemene, Kofele and Leku stations) indicates a total overestimation of rainfall intensity for events correctly detected by the CHIRPS, while a negative value (during the wet period at the Belela, Kokossa, and Wondo_genet stations) implies underestimated intensities by the CHIRP product. However, the higher negative values of MB in the V2 estimates than in the V0 SRP estimates indicate that the CHIRP product has fewer missed events than does the CHIRPS product.

At this stage of the procedure, it is sufficient to check the percent of volume error (PVE) and RMSE to determine the importance of resampling. Overall, the assessment using descriptive statistics (Table 6) illustrates that bilinear interpolation is a more efficient resampling method. Nevertheless, in terms of the PVE performance, at some stations (i.e., bolded in Table 5), the NN method outperforms the BL method. The assessment of bias decomposition methods (Annexes 6 & 7) illustrates that the bilinear interpolation method is a more efficient method of spatial disaggregation. In conclusion, the performances before and after resampling did not show exaggerated changes. This is attributed to the fact that the CHIRPS algorithm also utilizes the information recorded at ground stations. Nevertheless, (Gebremedhin et al., 2021) reported that CHIRPS products can be corrected using ground station records not used in the original product. Refer to the annex section for details on the results of bias decomposition performance measures.

The comparison of the resampling outputs clearly shows that the bilinear resampling technique has better performance on the CHIRPS product. The bilinear resampling technique has also been proven to perform better in upper Takeze watersheds (Gebremedhin et al., 2021). The total depth of missed events in the wet period is greater than that in the dry period for the CHIRPS product, while the reverse is true for the CHIRP product.

Resampling procedures are important for reducing the spatial mismatch between grid observations and point observations at ground stations. However, resampling methods need to be selected properly by taking into consideration the pros and cons of available methods in line with the objective of the study. In this particular study, bilinear resampling performed better than the nearest neighbor resampling technique, which is similar to the study conducted by (Gebremedhin et al., 2021).

4.1.2. Bias correction

The bias correction technique utilized in this study was parametric empirical quantile mapping, which requires determining the best known distribution to fit the gauge data and SRP. For similar satellite rainfall products and control observations (gauge observations), the cumulative distribution function might vary depending on the temporal resolution, geographical location, season, rainfall formation, and other factors (Ma et al., 2019). Accordingly, one should consider the applicability of known distribution functions for the selected colocated data pairs (Soo et al., 2020), which are different functions for different datasets. Hence, because of the temporal variability in rainfall, the distributions were fitted on a monthly basis.

Accordingly, in good distribution fit tests, the chi-square and Kolmogorov–Smirnov tests showed that the exponential distribution was a good fit for both the resampled CHIRPS product and ground observations. The Kolmogorov–Smirnov test was assessed against the critical value at the 5% significance level. The bias correction (Eqn. 5, 6, & 7) at available grid points was also tested for performance.

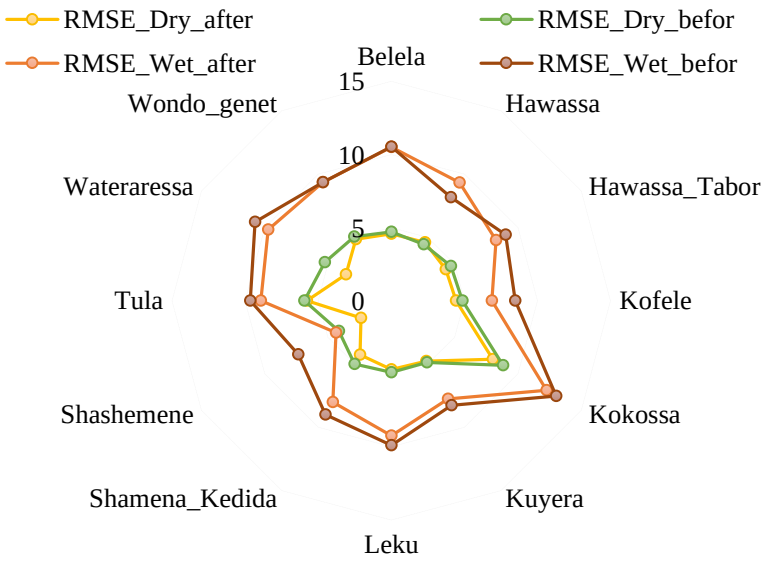


Figure 10 CHIRPS RMSE test results for dry and wet periods; before indicates the result after resampling, & after indicates the result after bias correction

Compared to the results acquired after the resampling stage (Annex 6), some stations (Shahemene, Kokossa, Belela, Hawassa, & Wondo_genet) exhibited poor performance, while the overall results suggested better performance and most definitely improved results compared to the raw CHIRPS dataset. According to the RMSE performance test results (Figure 10), the performance clearly improved for all stations except the Belela, Hawassa, and Wondo_genet stations.

A limitation of this process (i.e., specific to this study) is that no bias correction was performed to correct any rainfall event mismatches. Hence, one can clearly see that the bias correction resulted in poor performance of hit bias and miss bias in comparison to the resampled product. With the aim of correcting

grid points with no ground observations, bias correction using interpolated distribution parameters was tested using three hidden stations (Shashemene, Wateraressa, & Tula stations); the results indicate (Table 7) that the pattern was well captured, but the volume was underestimated except for the dry period at Tula station.

Table 7 Performance results of CHIRPS spatial bias correction using interpolated distribution parameters

	RMSE(mm)		PVE (%)		HB(mm)		MB(mm)	
Station	Dry period	Wet period	Dry period	Wet period	Dry period	Wet period	Dry period	Wet period
ShA	2.87	6.59	-75.09	-111.24	-863.50	-5010.20	704.30	2551.90
TU	4.02	7.80	5.57	-37.39	144.60	-3278.70	1717.60	4486.20
WT	3.53	8.24	-30.11	-21.17	-599.70	-1966.50	1439.90	5415.50

Moreover, the comparison against bias correction using parameters generated by using ground stations (Figure 11) indicates that information can be transferred to unrepresented grid points. Hence, future works should consider transferring bias correction parameters to ungauged grid points rather than using the Thiessen polygon to represent grid points without ground observations.

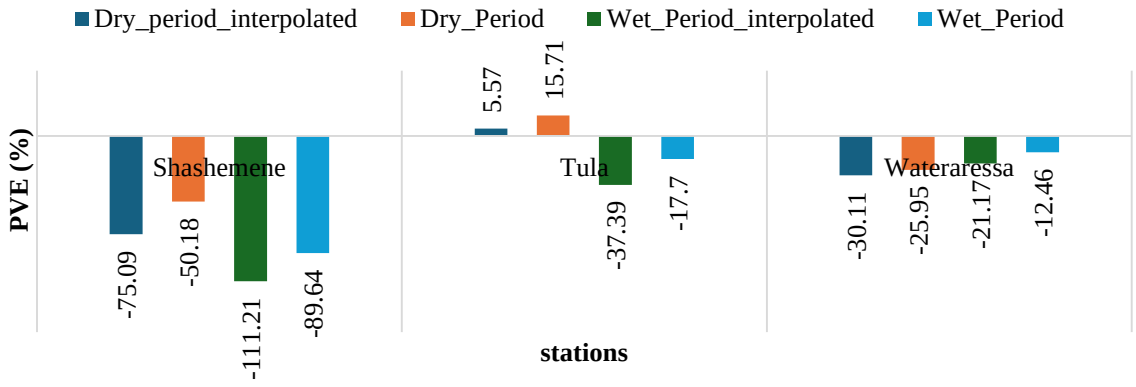


Figure 11 Performance test for CHIRPS using interpolated distribution parameters

Negative results (i.e., except for the RMSE) indicate that the estimated value underestimated the ground observation at that point, while positive results indicate overestimation. According to the results (Annex 9), most results suggest that pattern information can be transferred from grid points with ground observations to neighboring grid points with no ground observations. However, the results (except for the RMSE) at the Tula station are not in agreement with the results of the other two stations. This might be due to the limitation of the study, that no rainfall events were corrected.

Two conventionally used distributions were tested for bias by correcting the resampled CHIRP product. The results indicate that the exponential distribution (Table 8) best represents the wet period, while the dry period is well captured by the gamma distribution (Table 8). The results also indicated that false satellite rainfall detections can be eliminated or reduced by implementing moving window sampling techniques. The positive near zero results of FBI and PVE (Table 8) show very good agreement and a slight underestimation. However, compared to the NN performance (Table 8), the dry period is not well captured by the exponential distribution.

Table 8 At-station bias correction performance results for the exponential and gamma distributions

Station	Exponential distribution				Exponential distribution			
	PVE (%)		POD		PVE (%)		POD	
	Dry	Wet	Dry	Wet	Dry	Wet	Dry	Wet
BE	-42.76	3.03	0.77	0.98	0.01	11.94	0.56	0.89
HA	-47.69	3.16	0.86	0.98	2.48	11.92	0.63	0.89
HT	-39.39	0.91	0.85	0.98	1.72	11.73	0.62	0.88
KOF	-30.33	0.14	0.9	0.99	10.92	11.96	0.67	0.89
KOK	-18.53	0.83	0.83	0.99	30.75	13.85	0.54	0.88
KU	-61.87	0.14	0.85	0.99	-30.65	10.34	0.71	0.9
LE	-26.66	1.14	0.83	0.98	10.12	10.98	0.62	0.87
SHK	-61.55	0.13	0.86	0.99	-16.45	12.99	0.6	0.87
SHA	-40.96	7.02	0.83	0.94	14.42	17.54	0.54	0.84
TU	-33.89	0.98	0.82	0.97	7.48	10.56	0.57	0.89
WT	-33.5	2.87	0.86	0.97	7.33	14.27	0.57	0.86
WO	-56.04	1.92	0.85	0.98	-1.95	12.37	0.58	0.88

In contrast, comparing the HB and MB (Annexes 10 & 11) implies that missed events have not changed; rather, correctly captured events have increased in intensity. As an illustration, the hit bias at the Shamena Kedida station increased by 388.2 mm, while the miss bias increased by only 62.9 mm. The gamma distribution estimates mostly underestimate events except for those at the Kuyera and Shamena_Kedida stations during the dry period.

The RMSE improved slightly for bias corrections by the gamma distribution over the dry period, while the wet period performance improved using the exponential distribution (for detailed information, refer to the Annex section). Hence, combining these distributions, i.e., applying the gamma distribution for the dry period and the exponential distribution for the wet period, provides enhanced bias-corrected SREs.

Existing bias correction approaches are limited by their inability to correct nonrainfall events (Lehner et al., 2020). Similarly, underperformance was achieved by applying the parametric QM method without considering nonrainfall events. However, a significant improvement in the estimation was achieved at some stations, and when comparing the dry season to the wet season, the dry period performance improved.

Furthermore, a potential method of spatial bias correction for grid points for which no in situ records were available was tested. Encouraging results were achieved through this process, which signifies the importance of transferring bias correction parameters to grid points where ground measurements are not available.

4.1.3. Merged products

This study assessed two merged products, which helps to understand the advantages of each stage in the devised procedure. The first product was a merged rainfall product obtained by using available ground observations and resampled products. The validation procedure included cross-validation with 20 iterations, where in each iteration, two stations were randomly dropped, and the merging was performed using the remaining 10 stations. After merging, the performance test was performed at the hidden/dropped stations. The main reason to select two stations at a time was that it allows the remaining training data to be sufficient and to avoid completely unrepresented or distant observation to conduct interpolation.

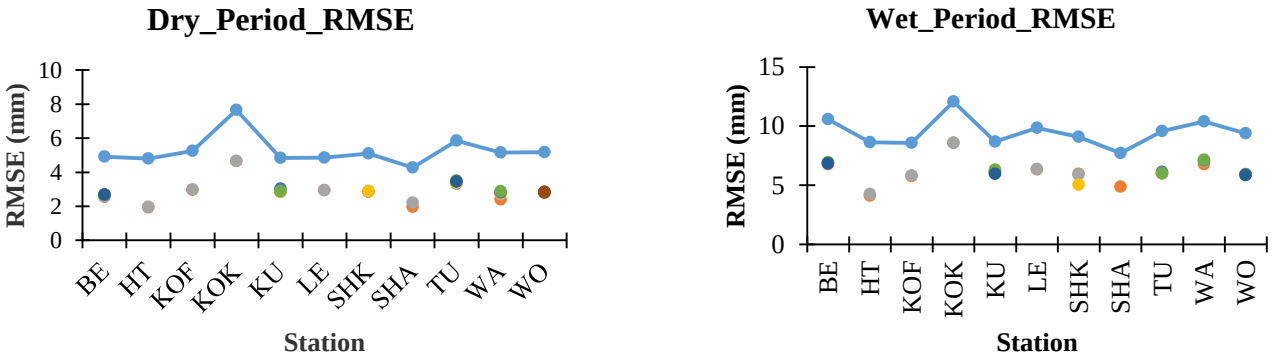


Figure 12 Conditionally merged RMSE results for all iterations (dots) compared to the RMSE results after resampling (solid line) for dry and wet periods

The comparison was performed against the test values of the bilinear resampling product. According to the RMSE performance test (Figure 12), on average, a 20% improvement in the RMSE was achieved. All other tests (not presented here) showed significant improvement. This suggests that simple conditional merging procedures are capable of reducing bias even for few ground records. Pignone et al. (2015) reported that this method can be effective in capturing rain gauge records and correcting SREs.

The merged output using the bias-corrected rainfall map and ground observations was tested using a similar cross-validation procedure over 20 iterations. The results indicate that all the iterations yield significantly improved performance. Comparing the dry period against the wet period, the performance of the dry period is much improved. However, this does not entirely indicate the ability of the procedure to correct dry period events better

than wet period events. This discrepancy could be partially attributed to the greater number of wet events in the wet period and the greater number of none rainfall events in the dry period.

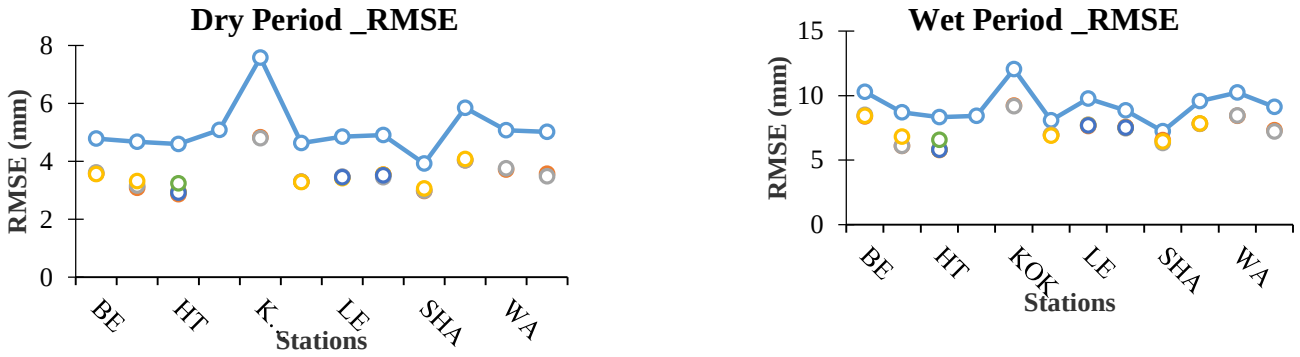


Figure 13 Bias-corrected and merged RMSE results for all iterations (dots) compared to the RMSE results after resampling (solid line) for dry and wet periods

Additionally, this study highlighted the significance of combining two or more observational sources to improve rainfall spatial estimates. Although the CHIRPS product is well known for its ability to retrieve ground rainfall measurements, a significant improvement in performance was achieved by combining the raw product with ground measurements. A similar deduction was made by different studies, such as (Gebremedhin et al., 2021, Navas et al., 2019). Furthermore, the simple conditional merging technique outperformed the devised fused bias correction method, where the parametric empirical QM method was followed by the conditional merging technique. However, this is due to the limitation of the study in not considering rainfall events in the bias correction process.

The after bias-corrected merged product of the CHIRP SRE showed significant improvement compared to the raw product. Specifically, the wet period statistical results were better than the dry period statistical results. Depending on the combination of coupled stations for validation, up to 70% PVE improvement for the wet period and up to 50% improvement were achieved at some stations (Shamena_Kedida, Wateraressa, and Hawassa_Tabor stations).

Table 9 Performance results for conditionally merged CHIRP SREs after bias correction

Stations	Dry Period		Wet Period	
	PVE (%)	POD	PVE (%)	POD
BE	-53.36	0.93	-13.04	0.99
HA	0.68	0.86	17.42	0.89
HT	-16.3	0.94	4.79	0.98
KOF	-10.17	0.91	-8.75	0.95
KOK	45.89	0.78	34.33	0.92
KU	-52.92	0.87	-17.77	0.8
LE	-17.64	0.95	-19.8	0.98
SHK	-17.64	0.87	-17.77	0.8
SHA	-81.45	0.9	-72.97	0.81
TU	9.78	0.9	-15.77	0.98
WT	-6.51	0.81	-5.84	0.94
WO	15.7	0.79	16.29	0.92

From the hit bias performance (Annex 12), one can observe that most correctly detected events are underestimated (i.e., negative values), except for those of Hawassa and Kokossa in the wet season. The above result is the most efficient at ensuring that a station performs out of 30 iterations by randomly choosing two hidden stations at a time. A comparison of the bias-corrected and merged CHIRP outputs (Table 9) with the resampled CHIRPS estimates (Table 6) indicates that several performance measures (i.e., MB, HB, RMSE and POD) (Figure 14) are significantly improved.

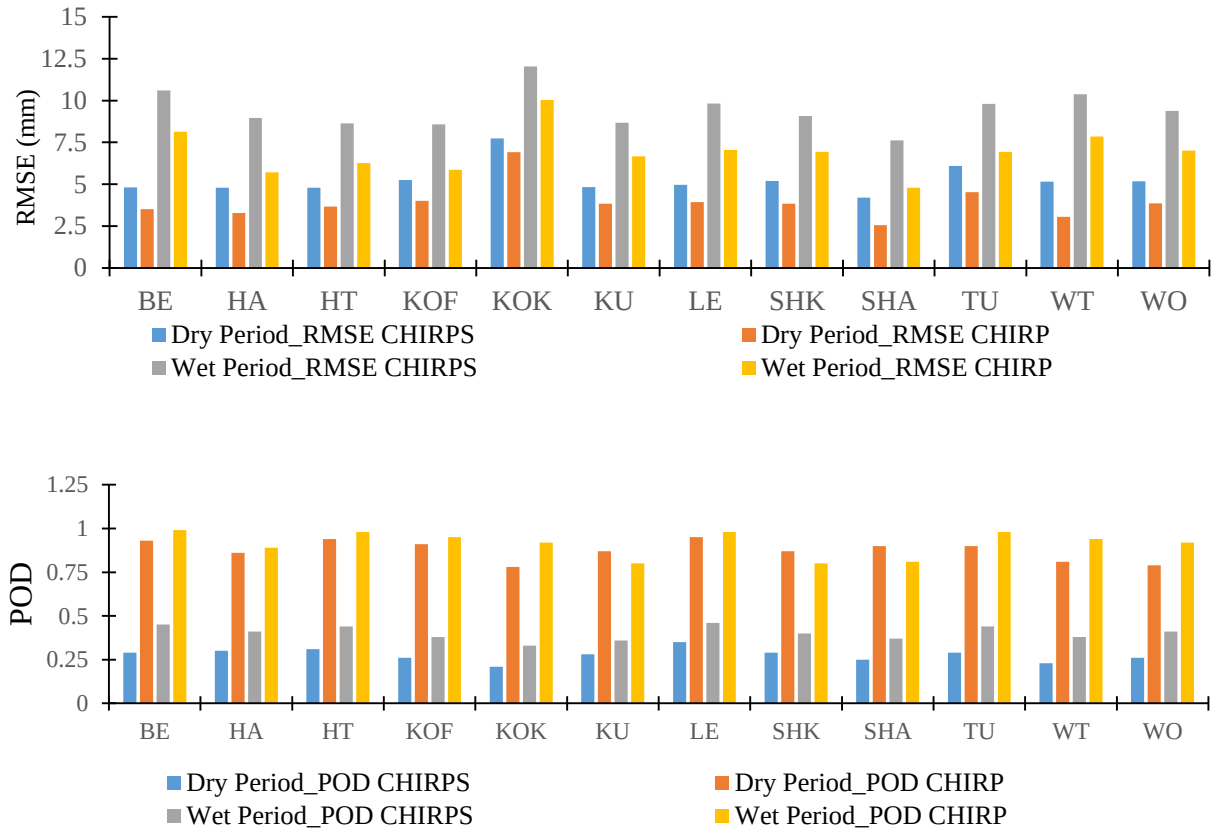


Figure 14 Comparison of bias-corrected and merged CHIRP with the resampled CHIRPS product

However, considering the error in the volume of the merged CHIRP product, the estimation was poor at the Kuyera station and Belela station for the dry period. Additionally, even though the errors in volume at the Kokossa and Shashemene stations improved, compared with the CHIRPS estimates, the overall performance was not satisfactory. In general, the results of the present study indicate that the devised procedure to correct CHIRP SREs can be more effective at reducing errors than the CHIRPS product.

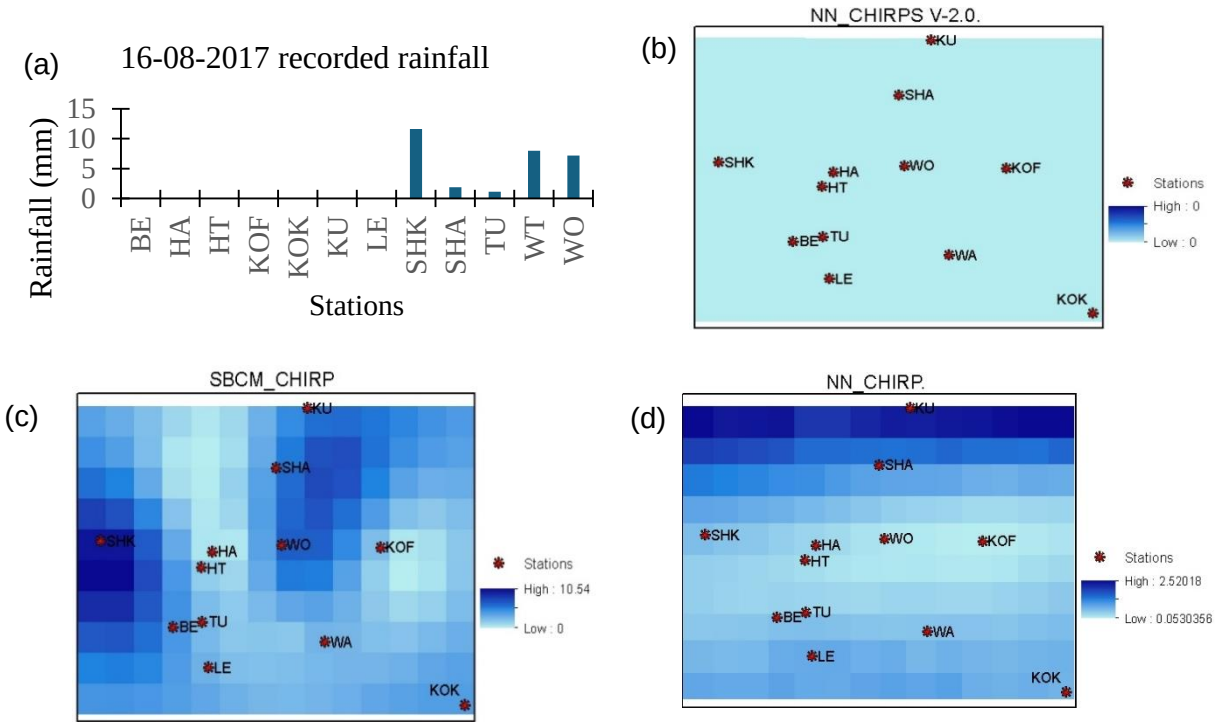


Figure 15 Spatial distribution of rainfall estimates on 16-08-2017 for (a) rainfall records at stations (b) resampled (NN) CHIRPS version 2.0 (c) spatially bias corrected and merged CHIRP (d) resampled (NN) CHIRP

The spatial distribution/variability of an event on the 16th of August/2017 (i.e., by assuming that gauge records (Figure 15-a) are the truth) illustrates that estimates by the resampled CHIRPS Version 2.0 (Figure 15-b) have totally missed the event. However, the estimates from the spatially bias-corrected and merged CHIRP products (Figure 15-c) captured the spatial coverage as well as the magnitudes at some stations. On the other hand, from the estimates of the resampled CHIRP (Figure 15-d), one can see that the northern part is wetter than the rest of the map, where the gauge records suggest heavier rainfall events in the western and central parts of the map. Hence, the proposed method is capable of correcting spatial patterns as well as event magnitudes.

4.1.4. Monthly SRE performance

Monthly estimates of both SPPs (i.e., CHIRP and CHIRPS) were assessed to investigate the impact of bias correction at different temporal resolutions. By aggregating the daily raw SRP and bias-corrected SREs at a monthly temporal resolution, the monthly estimates were compared using three statistical performance measures (Eqn. 8, 9, and 10). The performance tests were not conducted seasonally at this stage. The upgraded CHIRPS product has fairly good estimates at the originally incorporated gauging station (i.e., the Hawassa gauging station). Examining the error in volume over the study period, bias-corrected (BC) CHIRP estimates outperformed the remaining estimates at all stations except at the Hawassa, Kokossa, and

Wondo_genet stations. Additionally, BC CHIRP estimates underestimated at the Kuyera, Wondo_genet and Leku stations.

Compared with the resampled CHIRPS estimates, the spatially biased corrected and conditionally merged (SBCM) estimates showed slightly better correlations at all stations except for the estimates at the Watararessa, Shashemene, and Kuyera stations. Hence, this study highlights the importance of multistage bias correction methods for improving bias corrections for ungauged gridded SREs. This is supported by the relatively better result of SBCM estimates in the underrepresented western part of the watershed, which is represented only by the Shamena_Kedida station.

Table 10 Comparison of the raw CHIRP and CHIRPS products at collocated ground stations with bias correction (BC) and spatially bias correction and conditional merging (SBCM)

Station	RMSE (mm)				r^2				PVE (%)			
	BC CHIRP	Raw CHIRP	Raw CHIRPS	SBCM CHIRP	BC CHIRP	Raw CHIRP	Raw CHIRPS	SBCM CHIRP	BC CHIRP	Raw CHIRP	Raw CHIRPS	SBCM CHIRP
WA	96.28	105.63	100.34	100.59	0.3	0.16	0.23	0.22	-3.55	-16.94	-14.77	-6.08
TU	45.13	47.23	41.26	40.22	0.58	0.52	0.63	0.67	-6.96	-11.57	-10.38	-9.31
HA	38.26	39.74	33.75	32.23	0.67	0.62	0.73	0.83	-5.86	-1.78	3.17	17.12
HT	43.51	46.06	44.8	42.79	0.52	0.49	0.5	0.53	-6.99	-14.53	-9.65	-5.6
SHA	21.09	53.48	49.77	59.9	0.66	0.46	0.51	0.31	-2.74	-89.41	-78.87	-74.95
KO	33.37	43.81	36.42	39.24	0.7	0.5	0.65	0.65	-6.84	-2.9	-0.86	-9.37
WO	43.04	45.55	41.87	42.66	0.66	0.58	0.66	0.69	9.66	3.54	7.34	16.18
SHK	46.68	51.81	48.93	42.35	0.57	0.53	0.56	0.63	-11.63	-39.79	-33.45	-22.83
LE	41.69	49.93	47.13	42.35	0.61	0.55	0.61	0.7	11.48	-20.07	-18.81	-19.37
KU	37.39	42.66	41.28	53.79	0.59	0.5	0.53	0.43	20.89	-20.95	-19.23	-23.2
KO	76.97	105.39	103.16	93.75	0.48	0.36	0.4	0.49	-3.66	43.38	42.9	37.59
BE	54.5	71.38	71.36	64.01	0.68	0.45	0.44	0.53	-2.95	-30	-25.04	-18.3

In conclusion, this study has proven that the SBCM approach is capable of improving the quality of SREs. The method is expected to improve the SREs in other areas, given that all the assumptions and requirements of the method are fully met. Specifically, considering the decline of ground station data supplied to archives such as the Climate Hazards Group (CHG) (Funk et al., 2015), the SBCM technique is useful for including unused datasets available from different sources. According to Funk et al. (2015), the number of ground station records available for the development of CHIRPS 2.0 declined from 2004 for the year 2004 to 500 for the year after 2010 in Africa. Hence, the method can be implemented for areas with supplementary rainfall records that are not considered in the development of CHIRPs version two. Furthermore, this study suggested that the performance of the SBCM improved for areas with few rain gauge stations. Thus, the implementation of the method for underrepresented areas is highly encouraged in comparison to the use of readily available bias-corrected or merged SREs.

4.2. Assessment of Input Rainfall for Groundwater Recharge Estimation Uncertainty

Meteorological data analysis

A meteorological data analysis was conducted for the study duration, which was from 2005 to 2015. A pre-analysis of the obtained meteorological time series dataset indicated that the mean annual rainfall at the Hawassa, Wendo-Genet, Kofele, and Wateraressa stations was 954.54, 1069, 1074, and 1152 mm, respectively. The temperature records of the Hawassa station reveal that the mean daily maximum temperature at the station is more than the corresponding temperatures at the Kofele and Wateraressa stations, which are located within mountain ranges, by up to 10°C. The evapotranspiration calculated from the available temperature data (Table 11) shows mean annual evapotranspiration (Figure 16) of 1608.5, 721.5, and 953.5 mm at the Hawassa, Kofele, and Wateraressa stations, respectively.

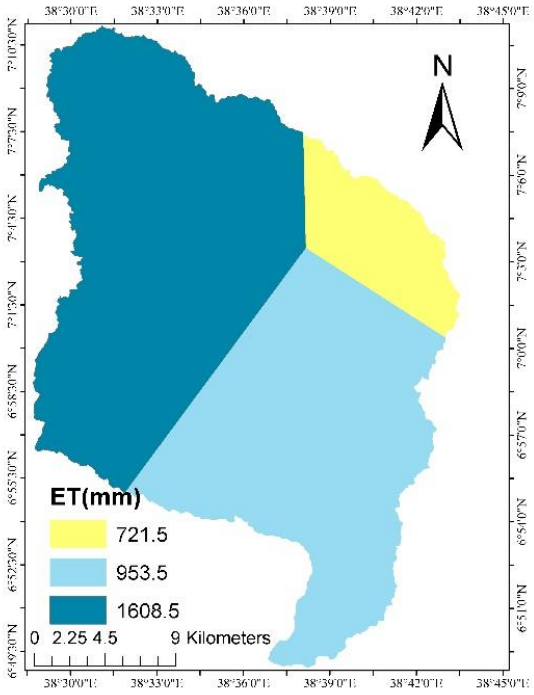


Figure 16 Mean annual evapotranspiration in the study area

The conventional methods used in previous studies (Lemlem, 2008, Yenehun et al., 2020, Abraham et al., 2022) to represent the spatial distribution of rainfall have mainly depended on the Thiessen polygon. These conventional methods are prone to error propagation in areas within a watershed that are sparsely represented by rain gauge stations, as is the case with the Tikur-Wuha watershed.

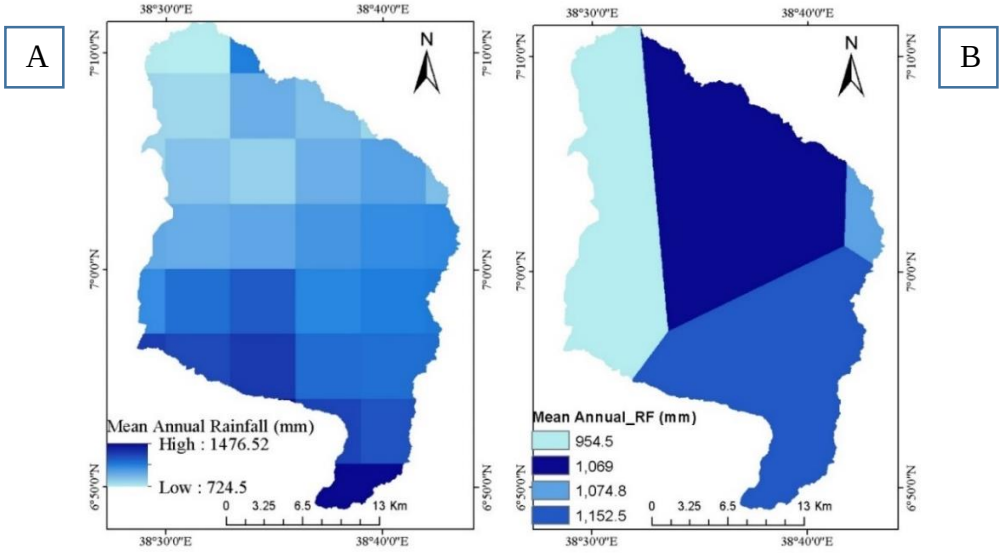


Figure 17 Mean annual rainfall distribution for the (B) M_1 and (A) M_2 model setups

As Figure 17 shows, the map prepared from the merged product obtained by combining CHIRP and rain gauge records (Figure 17(A)) is considerably different from the areal rainfall map created by employing the Thiessen polygon (Figure 17(B)) for the four rain gauge stations. This type of map discrepancy can have a significant impact on hydrological simulations. Improved spatial rainfall records or representations would significantly enhance the accuracy of recharge estimations, with recharge being directly proportional to rainfall. This improvement is indicated by the sharp edges of the recharge map (Figure 23), similar to the grids of the CHIRP-based SREs and the rainfall zones produced by the Thiessen polygon (Figure 17). The following sections provide additional details on the effects of each model input.

Table 11 Mean daily minimum and maximum temperature values and annual evapotranspiration values recorded at the Hawassa, Kofele, and Wateraressa rain gauge stations

Year	Minimum Temperature (°C)			Maximum Temperature (°C)			Annual Evapotranspiration (mm)		
	Hawassa	Kofele	Wateraressa	Hawassa	Kofele	Wateraressa	Hawassa	Kofele	Wateraressa
2005	12.81	7.63	8.32	27.57	19.74	19.47	1604.96	725.13	907.20
2006	13.58	7.09	8.80	27.32	19.44	19.18	1565.69	695.45	864.64
2007	13.02	6.36	9.02	27.13	19.46	20.66	1538.42	698.02	1034.58
2008	12.86	5.40	6.13	27.27	19.72	21.09	1566.10	726.99	1101.77
2009	13.38	4.66	4.65	28.24	20.14	19.03	1697.11	760.65	854.65
2010	14.21	4.37	5.73	27.04	19.61	18.10	1522.26	707.81	758.28
2011	13.53	3.04	5.74	27.70	19.95	17.61	1623.35	745.51	715.36
2012	13.37	1.73	8.74	27.99	19.88	20.01	1668.57	741.60	958.19
2013	13.00	1.92	9.93	27.57	19.26	21.13	1602.31	681.36	1088.86
2014	13.45	8.74	10.86	27.34	19.33	21.10	1565.37	685.94	1084.88
2015	14.34	8.68	11.10	28.54	20.25	21.37	1739.46	767.73	1123.93

4.2.1. Land Use/Cover Classification

How the sensitivity of recharge estimations can be influenced by land-use maps has been the subject of several previous studies (Abraham et al., 2022, Ampe et al., 2012), which have shown that careful consideration of land-use maps could enhance the sensitivity of recharge estimations. Hence, after considering the influence of the sensitivity of recharge estimations on land-use maps, a land-use map of the study area (Figure 18) was developed for the year 2014. The training and validation accuracies achieved for land-use classification were 0.965 and 0.864, respectively.

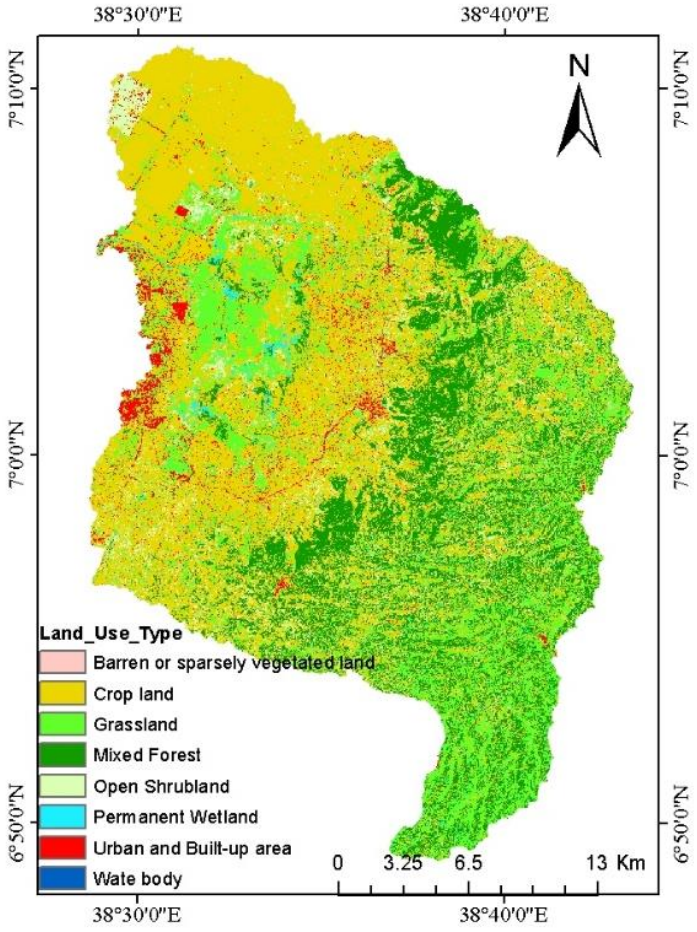


Figure 18 Land-use map of the Tikur-Wuha watershed prepared for the year 2014

As Table 12 indicates, cropland covers the largest proportion (46.72%) of the total area of the watershed, while 21.35% of this area is covered by mixed forest. According to the land classification performed by Mohammed et al. (2020), in 2017, 46.6% and 23.6% of the total area of the watershed was cultivated land and forest cover, respectively (Mohammed et al., 2020). Abraham et al. (2022) reported that in 2017, 58.62% and 25.5% of the watershed area was covered by cultivated land and forest, respectively (Abraham et al., 2022).

Table 12 Confusion matrix of each land-use class with its percentage coverage area

Land_Use_Type	Code	12	17	10	11	5	13	16	7	Percent
Crop land	**12**	91	0	1	0	0	1	0	0	**46.72**
Water body	**17**	0	11	0	1	0	0	0	0	0.005
Grassland	**10**	2	0	49	0	0	0	0	0	**19.14**
Permanent wetland	**11**	0	0	2	19	0	0	0	0	1.520
Mixed forest	**5**	0	0	0	1	37	0	0	0	**21.35**
Urban and built-up area	**13**	3	0	1	0	0	44	0	1	3.239
Barren or sparsely vegetated land	**16**	1	0	0	0	0	0	3	0	0.138
Open shrub land	**7**	1	0	1	0	0	0	0	31	7.885

4.2.2. Model Calibration

Model calibration was performed at the Dato village and Wosha stream gauge stations. A widely accepted practice is calibrating a given model using multiple gauge data. Stream gauge data are available at the outlet of the Tikur-Wuha River and its tributaries. Hence, in this study, model calibration was performed using the gauge records of the Tikur-Wuha River outlet and the data collected from a tributary located upstream of the Wosha subwatershed. Nevertheless, the NSE (0.28) of the calibration performed at the Dato village station was poor. Possible parameter combinations were explored to improve the Dato village station's calibration results. Additionally, the implementation of strategies such as removing peak values from the observed discharge time-series data and changing the simulation time periods was investigated. However, the maximum NSE obtained was 0.28. In contrast, the NSE of the calibration made at the hydrological station, located upstream of Wosha, was 0.56. Despite the calibration discrepancies, no attempt was made to transfer the calibration results obtained from the Wosha subwatershed gauge station to the Tikur-Wuha watershed gauge station.

The estimated long-term mean annual recharge obtained from the semi-distributed simulation model with parameters calibrated for the Wosha subwatershed was 218.29 mm. The long-term mean annual recharge obtained using the fully distributed simulation model with similar parameters ranged from 0.24 to 579.96 mm. Figure 6 shows a significant mismatch between the measured stream records obtained from the two discharge gauging stations. Another significant variation is the difference in the range of recharge obtained from model parameter calibration using the two discharge records.

The runoff in the study area increases as the recharge rate decreases (Figure 19) according to the calibrated model at Wosha station; the mean discharge is three times the mean discharge measured at the Dato village station. Ayenew and Tilahun (2008) argued that fault structures significantly enhance the recharge at the floor of a caldera (Ayenew and Tilahun, 2008). Another argument is that the contribution of the wetland to the watershed water balance cannot be disregarded when assessing the recharge (Abraham et al., 2022). In

depicting the effects of various land-use classes on water balance component modelling, implementing separate values for the local parameters of each land-use class would be helpful. The WetSpa model uses permanent wetland, a main land cover class, to calculate the local parameter values. Thus, the impact of wetlands on the simulated recharge was properly considered. The WetSpa model does not consider point recharge and/or wetland contributions to recharge as concentrated seepage through existing fault structures. This limitation can make recharge estimation challenging.

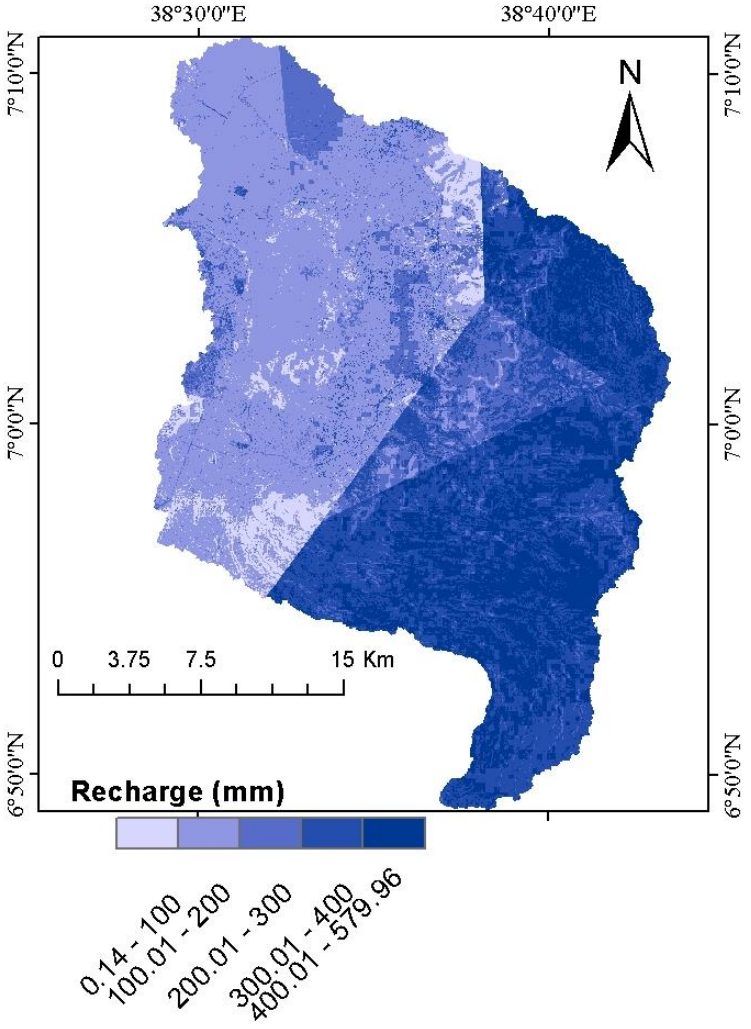


Figure 19 Long-term mean annual recharge in the study area obtained using Wosha station records

The discrepancies between the recharge magnitudes obtained from the two models calibrated using the Wosha and Dato-village records can be attributed to several factors. However, while the magnitude of recharge increases in an effort to match the discharge size at Dato village, it decreases as the flow rate at Wosha station is more accurately estimated. Because fault structures exist between the two recording stations, point recharges can therefore be considered the primary sources of recharge.

4.2.3. Recharge Uncertainty

Semi-distributed model simulation

As already explained, the M1 and M2 model setups differ in their input rainfall datasets. Thus, the land-use local parameters, soil hydraulic parameters, and meteorological input datasets used in each model run are the same except for the rainfall time-series data. In obtaining the merged rainfall dataset, the original cell size of the CHIRP-based SRE was considered for all rainfall regions. The resulting map had 37 raster cells ($0.05° \times 0.^{05°}$) with their rainfall records. Each grid cell center was assumed to have its station record, fed to the simulation model.

The M_1 model, which simulated 10,000 parameter combinations, produced a mean annual recharge in the range of 392.03–549.3 mm, whereas the semi-distributed M_2 model, which simulated 10,000 parameter combinations, produced a mean annual recharge in the range of 316.4–474.9 mm. The results indicate the impacts of input rainfall estimation uncertainties. The large difference between the minimum and maximum simulated recharge values indicates the uncertainty of the recharge estimates.

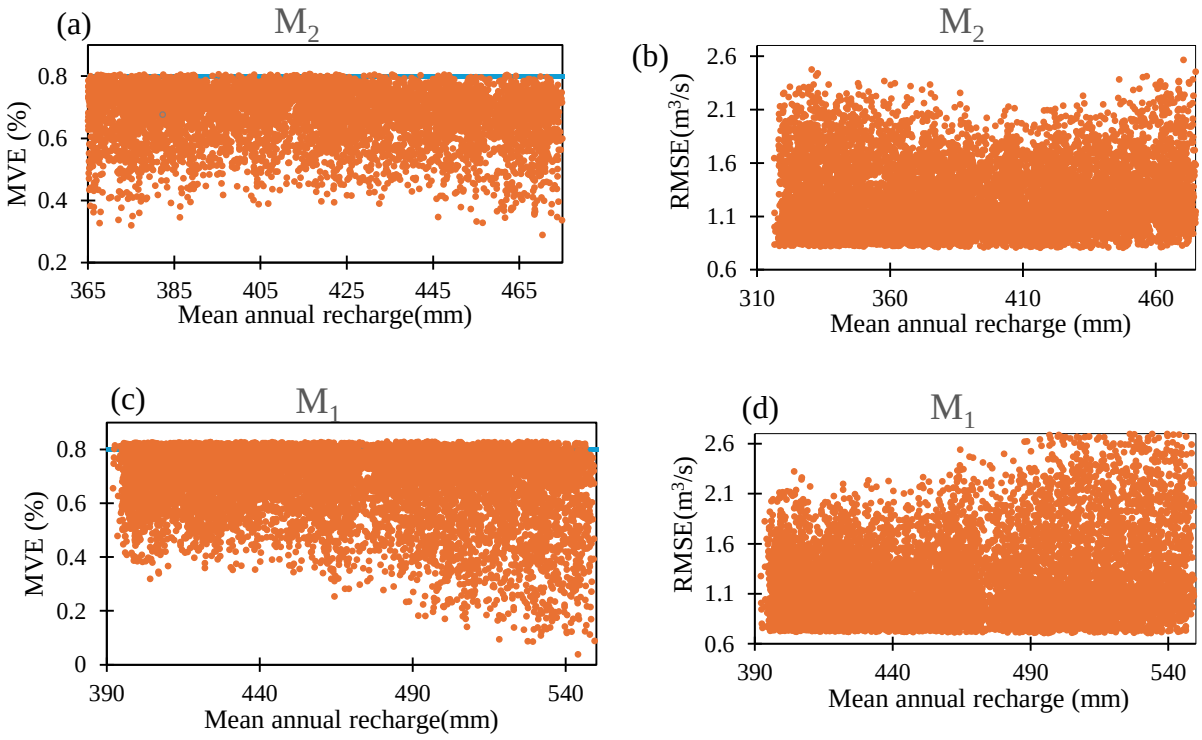


Figure 20 PVE and RMSE values obtained for 10,000 model simulations using the M_1 (c & d) and M_2 (a & b) model setups

The iteration results obtained from the M1 model setup demonstrated that both the RMSE (Figure 20(d)) and model volume error (MVE) (Figure 20(c)) increased as the mean annual recharge increased (Figure 20). The iteration results of the M2 model show that the RMSE (Figure 20(b)) increases and the MVE

(Figure 20(a)) decreases as the mean annual recharge increases; condensed values can be seen in the middle sections of the plots. The RMSE values obtained for 10,000 simulations performed using the M2 and M1 setups were between 0.8 to 2.57 m3/s and 0.8 to 3.27 m3/s, respectively.

The results obtained from both model setups indicated that the long-term mean annual recharge ranged from 370–430 mm. The RMSE values obtained from the M_1 model setup are densely distributed at low values of the long-term mean annual recharge. In contrast, the RMSE values obtained from the M_2 model setup are dispersed at the beginning and end of the X axis, depicting long-term mean annual recharge values. An exhaustive search of random parameter combinations that will ensure performance measure convergence, guided by a literature review and expert knowledge of the hydroclimatic characteristics of the watershed, has helped to explore the global ranges of each parameter. These predefined ranges of the parameter values were used at the start to generate possible parameter combinations randomly. The use of Monte Carlo simulation in conjunction with Latin hypercube sampling to generate possible parameter combinations helped to ensure that the parameters were properly distributed in the given parameter space.

The plot of the evapotranspiration coefficient (K-ep) against recharge (Figure 21) shows an inverse correlation between the mean annual recharge and K-ep; that is, the mean annual recharge decreases as K-ep increases. As K-ep increases, PVE also increases. This correlation indicates the importance of obtaining accurate evapotranspiration estimates. Thus, we can conclude that improved estimates of evapotranspiration can reduce the uncertainty of recharge estimations.

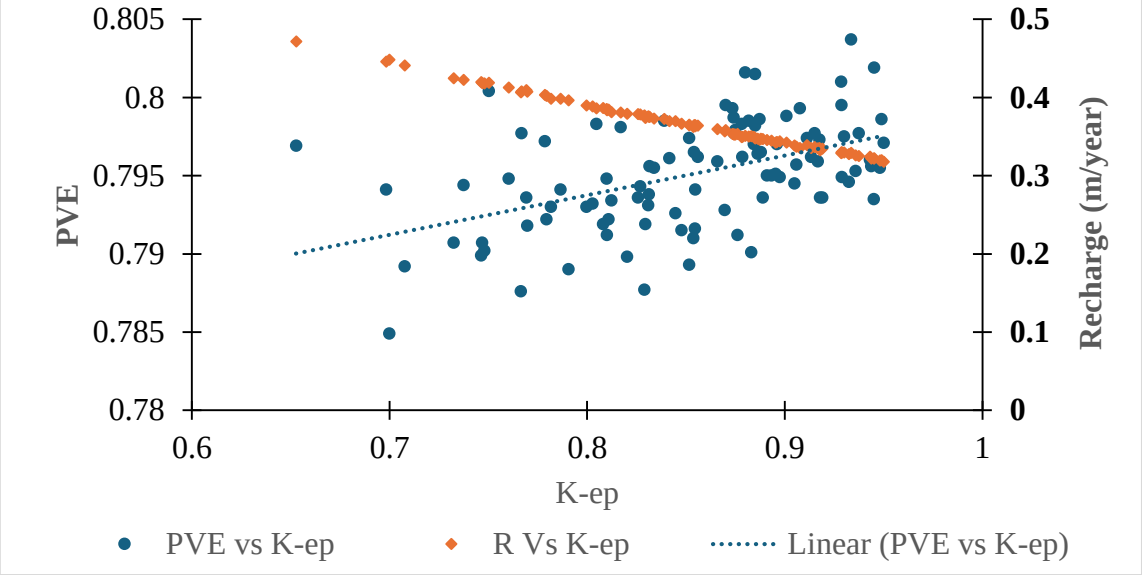


Figure 21 Sensitivity analysis between the evaporation coefficient (K-ep) and percentage volume error (PVE) of the estimated mean annual recharge (R) for the 100 best parameter combinations

Fully distributed simulation

The fully distributed simulation model available with the WetSpa extension generates the recharge accumulated in a targeted basin over the simulation period, taken as 10 years in this study. The average long-term recharge obtained from the 100 best spatial recharge maps produced by the M_1 model (Figure 22 (A)) is significantly different from that generated using the M_2 model. The model outputs depend on the input rainfall maps. Additionally, the highest and lowest long-term mean recharge are different for M_1 and M_2, where the recharge estimates of M_2 are greater than those of M_1. CHIRP-based SREs are noted for their overestimated rainfall values (Dinku et al., 2014, Dinku et al., 2011), which have been reflected in the highest recharge estimate produced by the M_2 model setup. According to Beyene et al. (2023), the raw CHIRP rainfall estimates show a false alarm ratio as high as 50%, indicating a noticeable increase in the depth of estimated rainfall values. Various bias correction techniques can be applied to reduce the inaccuracies present in volume estimations. A multistage bias correction strategy (Beyene et al., 2023) improved the PVE of CHIRP-based SREs by up to 70%. The overestimation can be decreased by using other available bias correction techniques. The two images shown below indicate the substantial spatial heterogeneity of the two standard deviation (STDEV) maps (Figure 22).

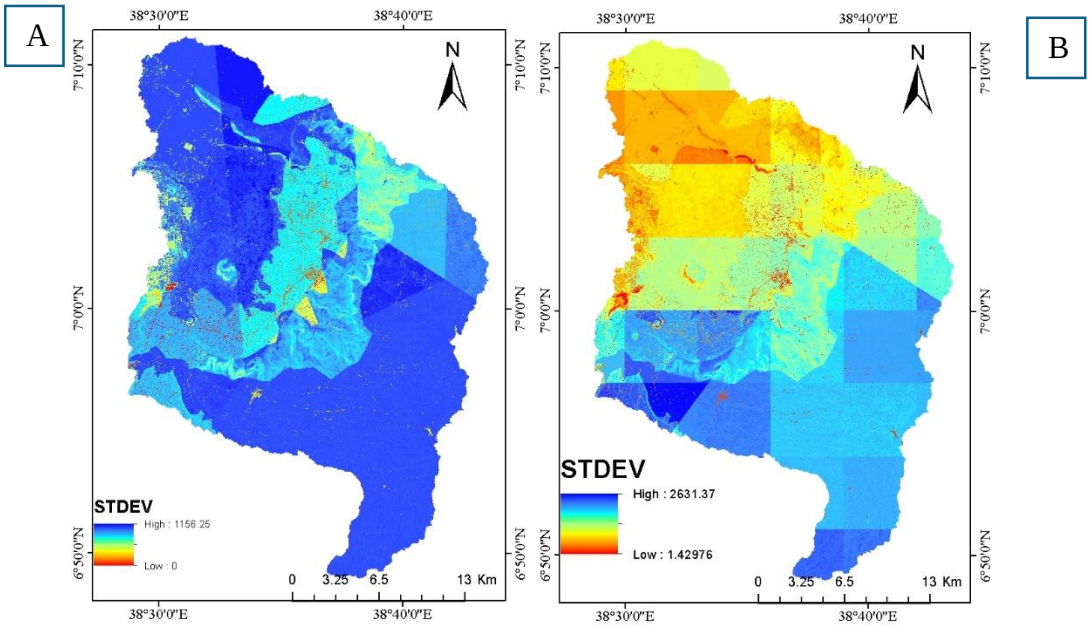


Figure 22Standard deviations of the 100 model simulations produced by the M1 (A) and M2 (B) model set-ups

Another strategy used to analyse the uncertainty in recharge estimations was to observe the STDEV of the selected 100 model realizations obtained from each model setup. A low STDEV indicates low uncertainty, while a high STDEV indicates high uncertainty in the estimations. The results show that merged SREs produce recharge maps with low uncertainty. The plots pertaining to the M_1 model setup shown on the left

(Figure 22(A)) indicate a maximum STDEV of 1156.25 mm, with most of the areas in the study area having high STDEV values. The plots pertaining to the M_2 model (Figure 22(B)) exhibited the highest STDEVs in a small part of the watershed. The M_1 model setup plots indicate that most watershed areas have high uncertainty values.

By dividing the long-term average recharge by 10, the number of years in the simulation period, the mean annual long-term recharge map of the study area can be obtained (Figure 23). From the maps, the impacts on recharge magnitudes and spatial distributions due to rainfall and evapotranspiration can be observed. Figure 23 shows a pattern similar to that of the Thiessen polygon map (Figure 16) developed using the evapotranspiration records obtained from monitoring stations. This pattern indicates that evapotranspiration has a significant impact on recharge estimation.

From the long-term mean annual recharge maps shown in Figure 23, the continuous impact of evapotranspiration can be observed. The western flat lands, dominated by crop land (Figure 18), of the watershed exhibited a low recharge magnitude (Figure 23 (A, B)), whereas the eastern mountain ranges exhibited high recharge magnitudes.

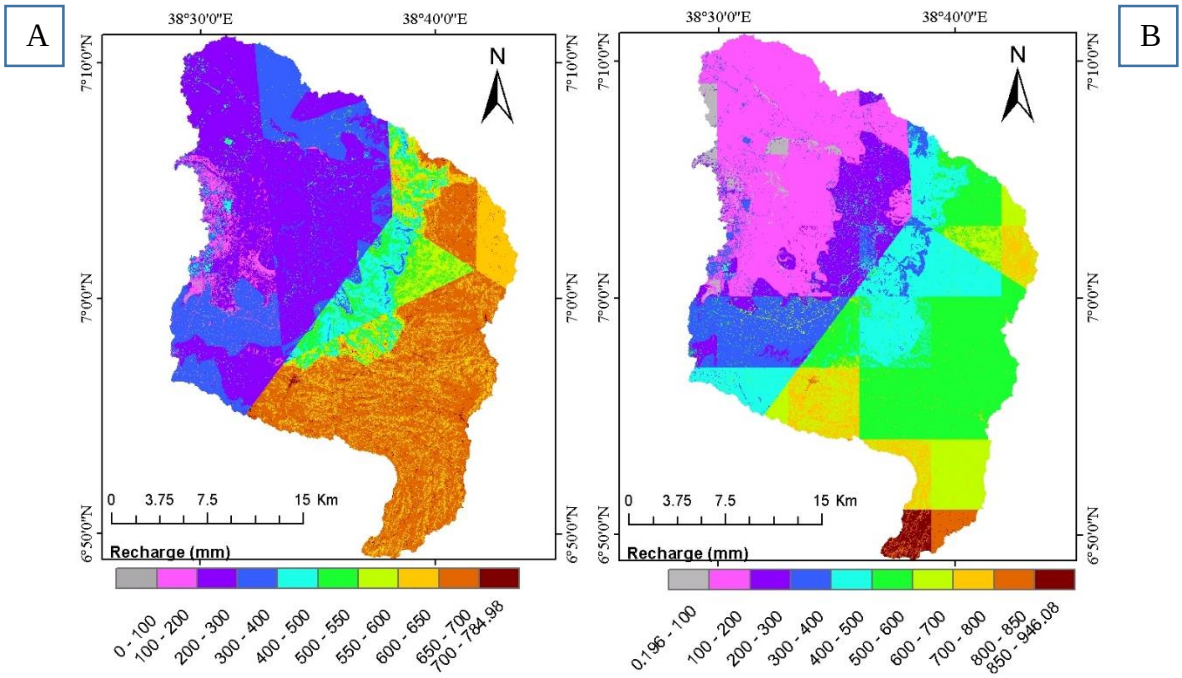


Figure 23 Long-term mean annual recharge map of the M_1 (A) and M_2 (B) model setups

According to previous studies where similar procedures were followed (Xie et al., 2018), model calibrations performed to determine possible parameter ranges can decrease the uncertainty in recharge estimations. However, this study relied on uncalibrated parameter ranges, which were determined based on the literature

and prior information related to the watershed. Thus, the output recharge map does not indicate the true or near-true values of the recharge in the watershed. Nonetheless, the results have illustrated the important role played by spatial rainfall distributions in recharge estimations.

The coefficient of variation (Cv) values obtained from the 100 best recharge maps produced by the M_1 model (Figure 24) indicated more consistent recharge estimations compared with the corresponding values obtained by the M_2 model. The highest uncertainty in recharge estimations was observed in the wetland-dominated part of the watershed. Abraham et al. (2022) also reported similar high recharge variability in the wetland-dominated part of the watershed (Abraham et al., 2022). Seventy percent of the watershed showed a Cv < 0.15 for the M_2 model setup (Figure 24(B)), while 90% of the watershed exhibited a Cv < 0.15 for the M_1 model setup (Figure 24(A)).

The recharge estimation sensitivity was studied by Abraham et al. (2022) using the calibrated soil water assessment tool (SWAT) tool SWAT-Cup, and the median recharge for 2017 was found to be in the range of 397 to 575 mm (Abraham et al., 2022). Despite the differences in the model setups, simulation periods, and model structures, significant variations can be observed between the recharge zone results of Abraham et al. (2022) and those of this study. According to Abraham et al. (2022), the wetland-dominated central part of the watershed exhibited a high median recharge, whereas this study showed that it had a low median recharge (Abraham et al., 2022). The estimated median recharge was low in the eastern mountainous parts of the watershed (Abraham et al., 2022). The semi-distributed simulation performed in this study using the M_1 model setup produced a mean annual recharge between 392.03 and 549.3 mm.

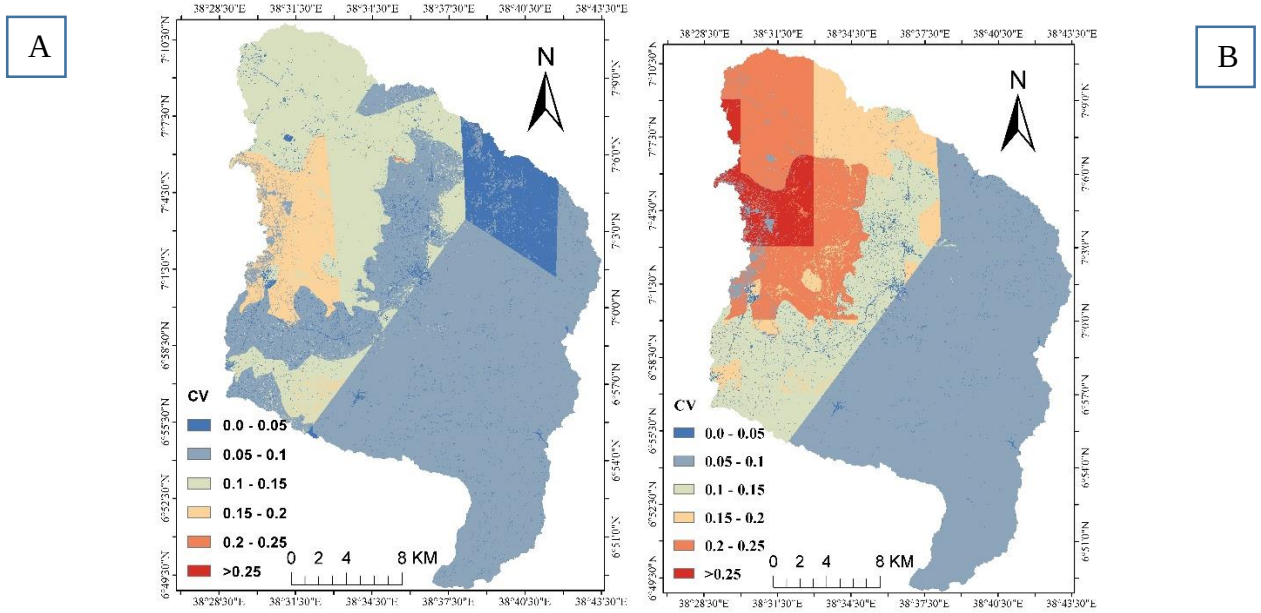


Figure 24 Coefficient of variation of the study area obtained from the (A) M_1 and (B) M_2 model setups

4.3. Evaluation of GLEAM potential evapotranspiration data for groundwater recharge estimation

4.3.1. Evapotranspiration data

Version 3.8a GLEAM reanalysis ET data resulted in four grids intersecting the boundaries of the Tikur Wuha watershed. The analysis of the data indicates that the annual average potential evaporation ranges from 894.65 mm to 1028.39 mm. Comparing the results with the potential evapotranspiration data produced using ground measurements of temperature and Enku's method (Eqn. 3), the estimates show a maximum annual evapotranspiration of 1608.5 mm at the Hawassa station and a minimum of 721.5 mm at the WT station. Furthermore, the annual evaporation for the resulting four grids of GLEAM is as tabulated below.

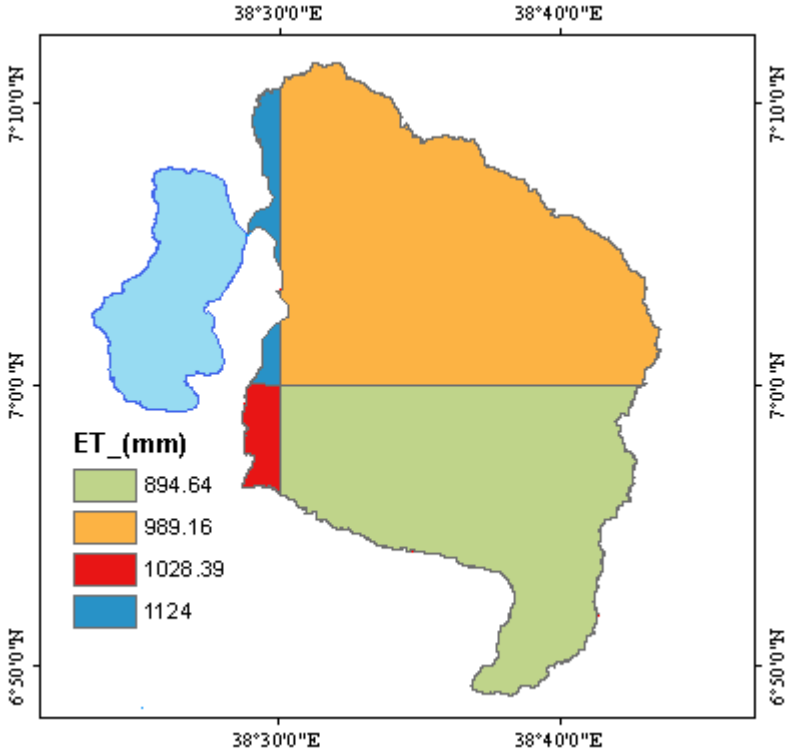


Figure 25 Annual average potential evapotranspiration (GLEAM 3.8a)

Similarly, the temperature data used for the estimation of evapotranspiration using Enku's method are missing. Additionally, the use of methods other than temperature-based methods is limited due to the unavailability of densely populated synoptic stations. Hence, the use of satellite-based reanalysis evapotranspiration data is essential for a better representation of evapotranspiration.

Table 13 GLEAM V3.8a annual potential evapotranspiration results for intersecting grids

Year	ST_1 (ET mm)	ST_2 (ET mm)	ST_3 (ET mm)	ST_4 (ET mm)
2005	1140.30	1008.40	1044.60	921.00
2006	1103.00	972.20	1019.60	877.70
2007	1129.20	992.60	1043.80	913.40
2008	1139.10	1004.80	1032.50	874.30
2009	1151.50	1004.00	1036.80	902.60
2010	1099.40	970.60	1006.40	927.50
2011	1133.40	998.10	1029.20	875.30
2012	1144.20	1001.30	1040.40	884.60
2013	1102.10	972.20	1015.50	888.60
2014	1103.00	967.40	1015.10	881.60
Average	**1124.52**	**989.16**	**1028.39**	**894.66**

4.3.2. Recharge uncertainty

Semi-distributed Simulation

The third simulation (M3) was executed by using the original grids of the GLEAM evapotranspiration data (i.e., 0.25° × 0.$^{25°}$) as the input Thiessen polygon. The center of the original grid cell represented a station, and 4 grids intersected the Tikur Wuha watershed. The only difference when comparing the third model setup (M3) with the model setup M2 is that the evapotranspiration data were replaced with the data extracted from GLEAM v3.8a. This means that all other data, such as the input rainfall, land use map, and soil map, are kept similar. To ensure the consistency of the iterations, 10,000 previously randomly generated global parameter combinations of the WetSpa model were used.

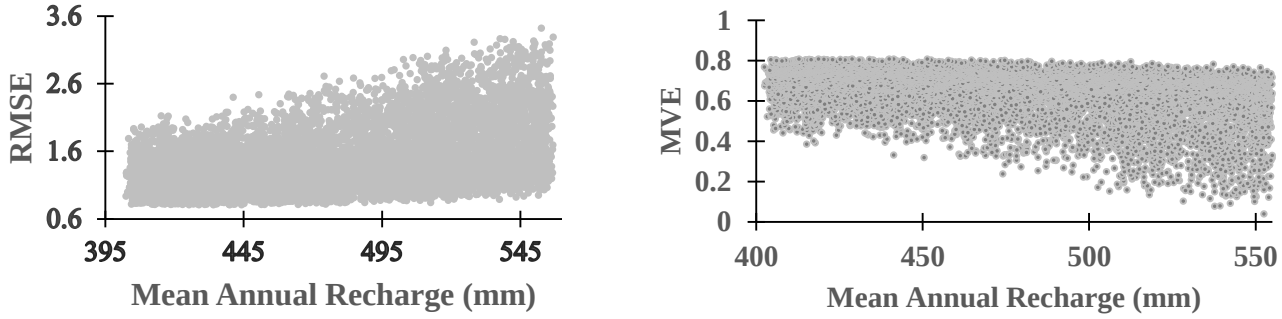


Figure 26 Resulting RMSE and PVE of 10,000 model realizations for the M3 model setup

The plot of the mean annual recharge versus the RMSE and MVE indicates that the model performs better for the mean annual recharge in the range of 400 mm to 450 mm. Further analysis of the 10000 simulations revealed that the average annual recharge reached 437.29 mm, with a standard deviation of 25.9 mm, when

the RMSE was < 0.9 m3/s. A cut-off value of 0.9 m^3/s was used to maintain sufficient model runs to conduct the uncertainty analysis.

Compared to the simulation results of the M2 model setup with a mean annual recharge of 375.27 mm and a standard deviation of 40.8 mm, the standard deviation of the simulation result of the M3 model setup significantly decreased. By comparing the standard deviation results, we can conclude that the use of GLEAM evapotranspiration reduced the uncertainty of the estimated recharge. However, the mean annual recharge increased from 375.27 mm to 437.29 mm.

Further comparison of the coefficient of variation for the selected model runs (Figure 27) of M2 and M3 revealed that the Cv for M2 was 0.108, whereas that for M3 was 0.059, which implies a 54% uncertainty reduction. In conclusion, the use of bias-corrected and merged CHIRP rainfall products and the use of GLEAM evapotranspiration data significantly reduced the uncertainty in groundwater recharge estimation for the Tikur-Wuha watershed.

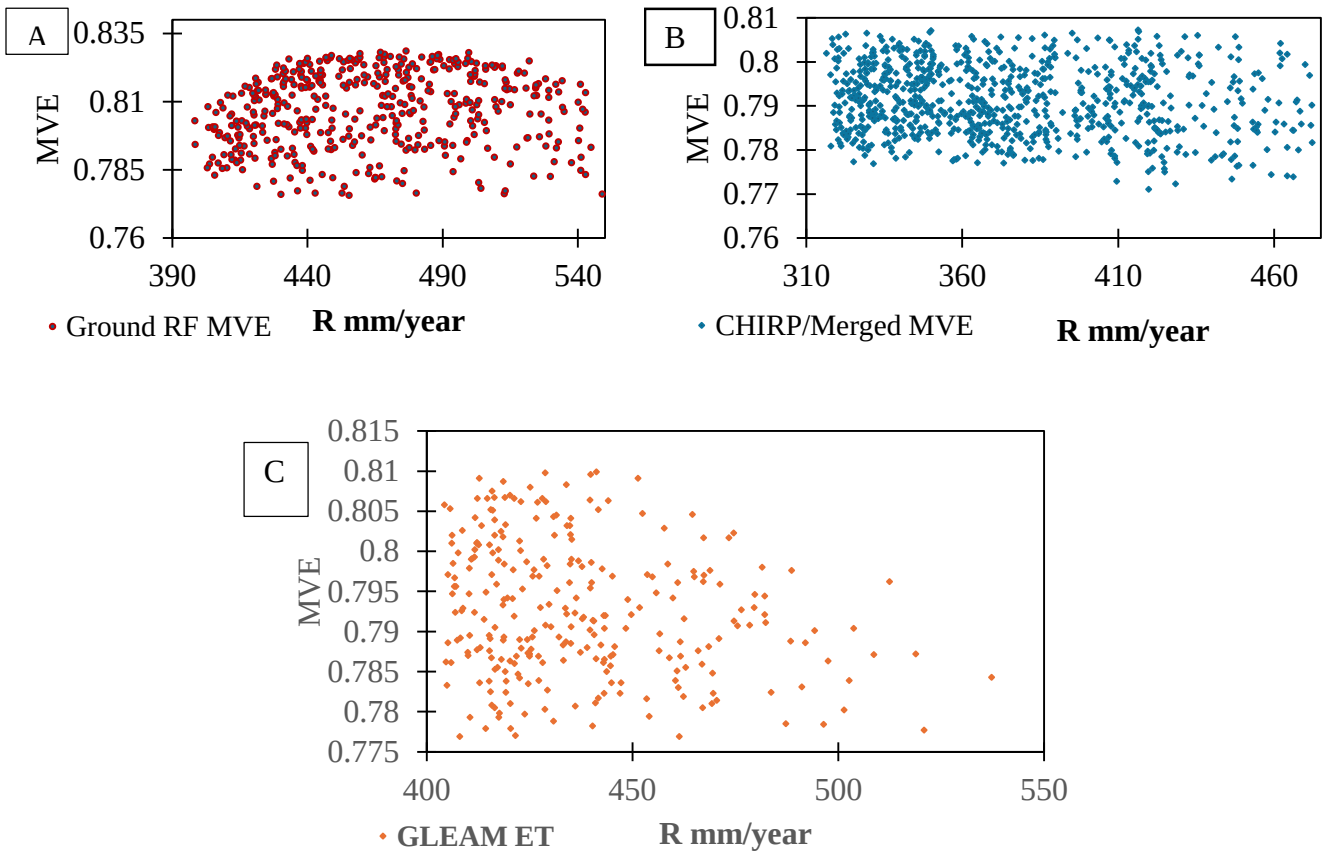


Figure 27 Comparison of mean annual recharge versus *MVE for selected model runs M1 (a), M2 (b) and M3 (c)*

CHAPTER FIVE

5.0. Conclusion and Recommendations

5.1. Conclusion

Accurate modelling of all components in the hydrological cycle is necessary for better planning and implementing sustainable water resource management. In view of the direct proportionality of recharge to rainfall records, this study was conducted to assess the uncertainty in recharge estimation due to input rainfall and evapotranspiration data. Recent technological developments have aided in the collection of precipitation data at both spatial and temporal scales. Additionally, the necessity of employing merging techniques to acquire the merits of different sources is increasing.

Hence, this study has the potential to produce accurate rainfall maps on a daily time scale by using a multistage procedure to correct the CHIRP and CHIRPS satellite rainfall products. Additionally, by evaluating the performance of each stage, the significance of each method devised in each stage was illustrated. The results demonstrated that the bilinear resampling technique performed better in the study area. Additionally, the objective resampling technique helps to reduce bias due to spatial mismatches between ground observations and selected grid estimates of SREs. The resampling outputs were bias corrected using the parametric empirical QM method. The results indicate that, when good fit distribution equations specific to a given dataset are used, bias correction of the raw satellite rainfall products can improve their performance compared to in situ observations. Further assessment of the transfer of bias correction parameters from grid points with in situ records to grid points with no ground observations was performed. The results illustrated that bias correction parameters can be effectively transferred from available grid points with control ground stations to other grids through interpolation.

Merging satellite rainfall products with ground observations is highly capable of preserving extreme events as well as preserving rainfall patterns. The results of the applied performance measures (i.e., seven) on the corrected and merged CHIRP SREs show that the percentage of detection (POD) and percent volume error (PVE) improved. Depending on the combination of coupled stations for validation, up to 70 and 50% PVE improvement was achieved at some stations for wet and dry periods, respectively. Moreover, over the sparsely populated western part of the Lake Hawassa basin, the bias-corrected and conditionally merged CHIRP SREs outperformed the estimates obtained by CHIRPS.

A physically based and distributed hydrological model of WetSpa was simulated with over 30,000 possible model parameter combinations. A total of 10,000 simulations each were executed using rain gauge-based

records, CHIRP SREs bias corrected and merged with available rain gauge records, and GLEAM ET estimates. All other meteorological inputs and physiographical input data were kept constant except for the input rainfall and evapotranspiration data. The results demonstrated that a better spatial representation of rainfall increases the accuracy of spatial recharge estimation. Additionally, the study showed that the spatial representation of evapotranspiration can have a significant impact on the estimation of long-term recharge distribution.

WetSpa model calibration with an NSE of 0.56 at the Wosha River gauge resulted in a long-term mean annual recharge of 218.29 mm for the Tikur-Wuha watershed. Seventy percent of the watershed showed a coefficient of variation (Cv) < 0.15 for M2, while 90% of the area exhibited a Cv < 0.15 for M1. The replacement of ET estimates of ground meteorological records with the GLEAM dataset reduced the Cv value by 54%. Hence, the study findings highlighted the importance of improving evapotranspiration data accuracy to reduce the uncertainty of recharge estimates.

5.2. Recommendations

These studies illustrated that applying rescaling techniques such as nearest neighbor and bilinear interpolation can improve the spatial estimates of rainfall. However, it is evident that this can introduce uncertainty and that the development of these methods for upscaling can increase the uncertainty. Hence, future studies are recommended to investigate this impact by considering more appropriate downscaling techniques and to further evaluate its implications for uncertainty propagation. It is also recommended that future studies focus on fusing techniques to assess the merits of different methods and provide improved bias correction procedures.

It is highly recommended to consider refining the base codes for the developed multistage bias correction algorithm when applying the method to other locations. Additionally, the performance achieved is limited to the study area and does not limit the procedure to other study areas. Furthermore, the study is limited in its ability to bias-correct no rainfall frequency. Hence, future works are recommended to investigate the performance of the algorithm and refine the code by considering the no rainfall distribution.

The study was limited in considering point recharges through faults, which are scattered throughout the watershed. Hence, the resulting recharge might not reflect the actual recharge of the watershed. Additionally, the study was limited to considering uncontrolled irrigation water use and total recharge from irrigation fields scattered across the watershed. Hence, future works should consider accounting for the contribution of irrigation water input to the total recharge of watersheds. Furthermore, better point recharge estimations and other spatial evapotranspiration estimations should be considered in future works.

References

ABDURAHMAN, S. G. & ZEWDIE, M. 2018. Fluoride ion and total dissolved solid distribution in Ethiopian Rift valley: The case of Hawassa city aquifer. *Journal of Hydrology: Regional Studies,* 19**,** 240-249.

ABIYE, T. A. 2008. Environmental resources and recent impacts in the Awassa collapsed caldera, Main Ethiopian Rift. *Quaternary international,* 189**,** 152-162.

ABRAHAM, T. & MULUNEH, A. 2022a. Assessing impacts of future climate on the crop water requirement and growth period. A case of Lake Hawassa watershed, Ethiopia.

ABRAHAM, T. & MULUNEH, A. 2022b. Quantifying Impacts of Future Climate on the Crop Water Requirement, Growth Period, and Drought on the Agricultural Watershed, in Ethiopia. *Air, Soil and Water Research,* 15**,** 11786221221135151.

ABRAHAM, T., MULUNEH, A., GIRMA, R., HARTMANN, A. & TEKLEAB, S. 2022. Quantifying sensitivity of groundwater recharge to land use and land cover changes by improving model performance on the wetland dominated Tikur Wuha watershed, Ethiopia. *Water Cycle,* 3**,** 112-125.

ADOMAKO, D., MALOSZEWSKI, P., STUMPP, C., OSAE, S. & AKITI, T. 2010. Estimating groundwater recharge from water isotope (δ2H, δ18O) depth profiles in the Densu River basin, Ghana. *Hydrological Sciences Journal–Journal des Sciences Hydrologiques,* 55**,** 1405-1416.

AHMET, Ş., İRFAN, Y. & ERCAN, S. 2020. Determination of recharge, storage and flow characteristics of a karst aquifer using multi-method approaches (Kocaeli, Turkey). *Hydrogeology Journal,* 28**,** 2141-2157.

AJAMI, N., DUAN, Q. & SOROOSHIAN, S. An Integrated Bayesian Uncertainty Estimator: fusion of Input, Parameter and Model Structural Uncertainty Estimation in Hydrologic Prediction System. AGU Fall Meeting Abstracts, 2005. H53H-08.

AKURUGU, B. A., OBUOBIE, E., YIDANA, S. M., STISEN, S., SEIDENFADEN, I. K. & CHEGBELEH, L. P. 2022. Groundwater resources assessment in the Densu Basin: A review. *Journal of Hydrology: Regional Studies,* 40**,** 101017.

ALCALÁ, F. J. & CUSTODIO, E. 2014. Spatial average aquifer recharge through atmospheric chloride mass balance and its uncertainty in continental Spain. *Hydrological Processes,* 28**,** 218-236.

AMPE, E. M., VANHAMEL, I., SALVADORE, E., DAMS, J., BASHIR, I., DEMARCHI, L., CHAN, J. C.-W., SAHLI, H., CANTERS, F. & BATELAAN, O. 2012. Impact of urban land-cover classification on groundwater recharge uncertainty. *IEEE Journal of selected topics in applied earth observations and remote sensing,* 5**,** 1859-1867.

AYENEW, T. 2001. Numerical groundwater flow modeling of the Central Main Ethiopian Rift lakes basin. *SINET: Ethiopian Journal of Science,* 24**,** 167-184.

AYENEW, T. & BECHT, R. 2008. Comparative assessment of the water balance and hydrology of selected Ethiopian and Kenyan Rift Lakes. *Lakes & Reservoirs: Research & Management,* 13**,** 181-196.

AYENEW, T. & TILAHUN, N. 2008. Assessment of lake–groundwater interactions and anthropogenic stresses, using numerical groundwater flow model, for a Rift lake catchment in central Ethiopia. *Lakes & Reservoirs: Research & Management,* 13**,** 325-343.

BABOO, S. S. & DEVI, M. R. 2010. An analysis of different resampling methods in Coimbatore, District. *Global Journal of Computer Science and Technology*.

BAHREMAND, A. 2016. HESS Opinions: Advocating process modeling and de-emphasizing parameter estimation. *Hydrology and Earth System Sciences,* 20**,** 1433-1445.

BAHREMAND, A., AHMADYOUSEFI, S., SHEIKH, V. & KOMAKI, C. B. 2021. A parameter allocation approach for flow simulation using the WetSpa-Python model. *Hydrological Processes,* 35**,** e13992.

BALIN, D., LEE, H. & RODE, M. 2010. Is point uncertain rainfall likely to have a great impact on distributed complex hydrological modeling? *Water Resources Research,* 46.

BEAR, J., CHENG, A. H.-D., BEAR, J. J. & CHENG, H.-D. A. 2010. Modeling contaminant transport. *Modeling groundwater flow and contaminant transport***,** 341-523.

BELETE, M. D., DIEKKRÜGER, B. & ROEHRIG, J. 2016. Characterization of water level variability of the main ethiopian rift valley lakes. *Hydrology,* 3**,** 1.

BELETE, M. D., DIEKKRÜGER, B. & ROEHRIG, J. 2017. Linkage between water level dynamics and climate variability: the case of Lake Hawassa hydrology and ENSO phenomena. *Climate,* 5**,** 21.

BEREHANU, B., AZAGEGN, T., AYENEW, T. & MASETTI, M. 2017. Inter-basin groundwater transfer and multiple approach recharge estimation of the upper Awash aquifer system. *Journal of Geoscience and Environment Protection,* 5**,** 76.

BEVEN, K. 1997. TOPMODEL: a critique. *Hydrological processes,* 11**,** 1069-1085.

BEYENE, T. D., MOGES, M. A. & TILAHUN, S. A. Development of rainfall disaggregation model in the Awash River Basin, Ethiopia. International Conference on Advances of Science and Technology, 2018. Springer, 50-64.

BEYENE, T. D., ZIMALE, F. A., GEBREKRISTOS, S. T. & NEDAW, D. 2023. Evaluation of a multi-staged bias correction approach on CHIRP and CHIRPS rainfall product: a case study of the Lake Hawassa watershed. *Journal of Water and Climate Change,* 14**,** 1847-1867.

BEYENE, T. D., ZIMALE, F. A., GEBREKRISTOS, S. T. & NEDAW, D. 2024. Assessment of the impact of rainfall uncertainties on the groundwater recharge estimations of the Tikur-Wuha watershed, rift valley lakes basin, Ethiopia. *Heliyon,* 10.

BHATTI, H. A., RIENTJES, T., HAILE, A. T., HABIB, E. & VERHOEF, W. 2016. Evaluation of bias correction method for satellite-based rainfall data. *Sensors,* 16**,** 884.

BITEW, M. M. & GEBREMICHAEL, M. 2011. Assessment of satellite rainfall products for streamflow simulation in medium watersheds of the Ethiopian highlands. *Hydrology and Earth System Sciences,* 15**,** 1147.

BRANDSMA, T. & KÖNNEN, G. 2006. Application of nearest-neighbor resampling for homogenizing temperature records on a daily to sub-daily level. *International Journal of Climatology: A Journal of the Royal Meteorological Society,* 26**,** 75-89.

BRESCIANI, E., ORDENS, C. M., WERNER, A., BATELAAN, O., GUAN, H. & POST, V. 2014. Spatial variability of chloride deposition in a vegetated coastal area: Implications for groundwater recharge estimation. *Journal of Hydrology,* 519**,** 1177-1191.

CARTWRIGHT, I., CENDÓN, D., CURRELL, M. & MEREDITH, K. 2017. A review of radioactive isotopes and other residence time tracers in understanding groundwater recharge: Possibilities, challenges, and limitations. *Journal of Hydrology,* 555**,** 797-811.

CHENG, H., LI, Y. & SUN, J. 2018. Interval Double-Sided Fuzzy Chance-Constrained Programming Model for Water Resources Allocation. *Environmental Engineering Science,* 35**,** 525-544.

CONTRERAS, L. F., BROWN, E. T. & RUEST, M. 2018. Bayesian data analysis to quantify the uncertainty of intact rock strength. *Journal of Rock Mechanics and Geotechnical Engineering,* 10**,** 11-31.

CROSBIE, R. S., BINNING, P. & KALMA, J. D. 2005. A time series approach to inferring groundwater recharge using the water table fluctuation method. *Water Resources Research,* 41.

CROSBIE, R. S., DOBLE, R. C., TURNADGE, C. & TAYLOR, A. R. 2019. Constraining the magnitude and uncertainty of specific yield for use in the water table fluctuation method of estimating recharge. *Water Resources Research,* 55**,** 7343-7361.

CROSBIE, R. S., PEETERS, L. J., HERRON, N., MCVICAR, T. R. & HERR, A. 2018. Estimating groundwater recharge and its associated uncertainty: Use of regression kriging and the chloride mass balance method. *Journal of Hydrology,* 561**,** 1063-1080.

CUSTODIO, E. 2010. Estimation of aquifer recharge by means of atmospheric chloride depositon balance in the soil. *Contributions to science***,** 81-97.

CUTHBERT, M. 2010. An improved time series approach for estimating groundwater recharge from groundwater level fluctuations. *Water Resources Research,* 46.

DAVIES, P. & CROSBIE, R. 2018. Mapping the spatial distribution of chloride deposition across Australia. *Journal of Hydrology,* 561**,** 76-88.

DEGEN, J. 2016. *Impact of land cover and soil conservation on soil erosion rates in the Tikur Woha Catchment, Ethiopia.*

DEGIFE, A., WORKU, H., GIZAW, S. & LEGESSE, A. 2019. Land use land cover dynamics, its drivers and environmental implications in Lake Hawassa Watershed of Ethiopia. *Remote sensing applications: society and environment,* 14**,** 178-190.

DELIN, G. N., HEALY, R. W., LORENZ, D. L. & NIMMO, J. R. 2007. Comparison of local-to regional-scale estimates of ground-water recharge in Minnesota, USA. *Journal of Hydrology,* 334**,** 231-249.

DELOTTIER, H., PRYET, A., LEMIEUX, J.-M. & DUPUY, A. 2018. Estimating groundwater recharge uncertainty from joint application of an aquifer test and the water-table fluctuation method. *Hydrogeology Journal,* 26**,** 2495-2505.

DENG, Z., PRIESTLEY, S. C., GUAN, H., LOVE, A. J. & SIMMONS, C. T. 2013. Canopy enhanced chloride deposition in coastal South Australia and its application for the chloride mass balance method. *Journal of hydrology,* 497**,** 62-70.

DILE, Y. T., TEKLEAB, S., AYANA, E. K., GEBREHIWOT, S. G., WORQLUL, A. W., BAYABIL, H. K., YIMAM, Y. T., TILAHUN, S. A., DAGGUPATI, P. & KARLBERG, L. 2018. Advances in water resources research in the Upper Blue Nile basin and the way forward: A review. *Journal of Hydrology,* 560**,** 407-423.

DINKU, T., CECCATO, P. & CONNOR, S. J. 2011. Challenges of satellite rainfall estimation over mountainous and arid parts of east Africa. *International journal of remote sensing,* 32**,** 5965-5979.

DINKU, T., HAILEMARIAM, K., MAIDMENT, R., TARNAVSKY, E. & CONNOR, S. 2014. Combined use of satellite estimates and rain gauge observations to generate high-quality historical rainfall time series over Ethiopia. *International Journal of Climatology,* 34**,** 2489-2504.

DOHERTY, J. E., HUNT, R. J. & TONKIN, M. J. 2010. Approaches to highly parameterized inversion: A guide to using PEST for model-parameter and predictive-uncertainty analysis. *US Geological Survey Scientific Investigations Report,* 5211**,** 71.

DUAN, Q., AJAMI, N. K., GAO, X. & SOROOSHIAN, S. 2007. Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Advances in Water Resources,* 30**,** 1371-1386.

DUBOVIK, O., SCHUSTER, G. L., XU, F., HU, Y., BÖSCH, H., LANDGRAF, J. & LI, Z. 2021. Grand challenges in satellite remote sensing. *Frontiers in Remote Sensing***,** 1.

EBRAHIM, G. Y., LAUTZE, J. F. & VILLHOLTH, K. G. 2020. Managed aquifer recharge in Africa: Taking stock and looking forward. *Water,* 12**,** 1844.

EILERS, A. 2018. *Deep groundwater characterisation and recharge estimation in the Verlorenvlei catchment.* Stellenbosch: Stellenbosch University.

ENKU, T. & MELESSE, A. M. 2014. A simple temperature method for the estimation of evapotranspiration. *Hydrological Processes,* 28**,** 2945-2960.

ESHETE, T. 2009. *Spatial analysis of erosion and land degradation leading to environmental stress: the Case of Lake Hawassa Catchment.* Doctoral dissertation, Addis Ababa University.

FANG, K., KIFER, D., LAWSON, K. & SHEN, C. 2020. Evaluating the potential and challenges of an uncertainty quantification method for long short-term memory models for soil moisture predictions. *Water Resources Research,* 56**,** e2020WR028095.

FAYER, M. J. 2000. UNSAT-H version 3.0: Unsaturated soil water and heat flow model theory, user manual, and examples. Pacific Northwest National Lab.(PNNL), Richland, WA (United States).

FENG, J., LIU, R., CHEN, P., YUAN, S., ZHAO, D., ZHANG, J. & ZHENG, Z. 2015. Degradation of aqueous 3, 4-dichloroaniline by a novel dielectric barrier discharge plasma reactor. *Environmental Science and Pollution Research,* 22**,** 4447-4459.

FRAGOSO, T. M., BERTOLI, W. & LOUZADA, F. 2018. Bayesian model averaging: A systematic review and conceptual classification. *International Statistical Review,* 86**,** 1-28.

FUNK, C., PETERSON, P., LANDSFELD, M., PEDREROS, D., VERDIN, J., SHUKLA, S., HUSAK, G., ROWLAND, J., HARRISON, L. & HOELL, A. 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific data,* 2**,** 1-21.

GAL, Y. & GHAHRAMANI, Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. international conference on machine learning, 2016. PMLR, 1050-1059.

GALLART, F., LATRON, J., LLORENS, P. & BEVEN, K. 2007. Using internal catchment information to reduce the uncertainty of discharge and baseflow predictions. *Advances in Water Resources,* 30**,** 808-823.

GDAL/OGR CONTRIBUTORS, W. 2020. GDAL/OGR geospatial data abstraction software library. *Open Source Geospatial Foundation*.

GEBERE, S. B., ALAMIREW, T., MERKEL, B. J. & MELESSE, A. M. 2015. Performance of high resolution satellite rainfall products over data scarce parts of Eastern Ethiopia. *Remote Sensing,* 7**,** 11639-11663.

GEBREMEDHIN, M. A., LUBCZYNSKI, M. W., MAATHUIS, B. H. & TEKA, D. 2021. Novel approach to integrate daily satellite rainfall with in-situ rainfall, Upper Tekeze Basin, Ethiopia. *Atmospheric Research,* 248**,** 105135.

GILLIES, S. 2013. The shapely user manual. *URL https://pypi. org/project/Shapely*.

GILLIES, S. 2019. rasterio Documentation. *MapBox, July,* 23.

GORELICK, S. M. & ZHENG, C. 2015. Global change and the groundwater management challenge. *Water Resources Research,* 51**,** 3031-3051.

GOSHIME, D. 2020. *Integration of satellite and ground-based rainfall data for water resources assessment in Central Rift Valley Lakes Basin, Ethiopia.* CY Cergy Paris Université.

GUENZI, D., FRATIANNI, S., BORASO, R. & CREMONINI, R. 2017. CondMerg: an open source implementation in R language of conditional merging for weather radars and rain gauges observations. *Earth Science Informatics,* 10**,** 127-135.

GUMUŁA-KAWĘCKA, A., JAWORSKA-SZULC, B., SZYMKIEWICZ, A., GORCZEWSKA-LANGNER, W., PRUSZKOWSKA-CACERES, M., ANGULO-JARAMILLO, R. & ŠIMŮNEK, J. 2022. Estimation of groundwater recharge in a shallow sandy aquifer using unsaturated zone modeling and water table fluctuation method. *Journal of Hydrology,* 605**,** 127283.

GUPTA, H. V., CLARK, M. P., VRUGT, J. A., ABRAMOWITZ, G. & YE, M. 2012. Towards a comprehensive assessment of model structural adequacy. *Water Resources Research,* 48.

HARRIS, C. R., MILLMAN, K. J., VAN DER WALT, S. J., GOMMERS, R., VIRTANEN, P., COURNAPEAU, D., WIESER, E., TAYLOR, J., BERG, S. & SMITH, N. J. 2020. Array programming with NumPy. *Nature,* 585**,** 357-362.

HASAN, S. H., AL-HAMEEDAWI, A. N. & ISMAEL, H. Supervised Classification Model Using Google Earth Engine Development Environment for Wasit Governorate. IOP Conference Series: Earth and Environmental Science, 2022. IOP Publishing, 012051.

HE, S., YU, K., TANG, Z., YAN, Y. & ZHANG, F. 2022. Impacts of parameter uncertainty on baseflow separation by a two-parameter recursive digital filter. *Hydrological Processes,* 36**,** e14512.

HEALY, R. W. 2010. *Estimating groundwater recharge*, Cambridge University Press.

HEALY, R. W. & COOK, P. G. 2002. Using groundwater levels to estimate recharge. *Hydrogeology journal,* 10**,** 91-109.

HEO, J.-H., AHN, H., SHIN, J.-Y., KJELDSEN, T. R. & JEONG, C. 2019. Probability distributions for a quantile mapping technique for a bias correction of precipitation data: A case study to precipitation data under climate change. *Water,* 11**,** 1475.

HEPPNER, C. S. & NIMMO, J. R. 2005. *A computer program for predicting recharge with a master recession curve,* US Geological Survey Menlo Park, CA, USA.

HEPPNER, C. S., NIMMO, J. R., FOLMAR, G. J., GBUREK, W. J. & RISSER, D. W. 2007. Multiple-methods investigation of recharge at a humid-region fractured rock site, Pennsylvania, USA. *Hydrogeology Journal,* 15**,** 915-927.

HØJBERG, A. & REFSGAARD, J. 2005. Model uncertainty–parameter uncertainty versus conceptual models. *Water Science and Technology,* 52**,** 177-186.

HU, Q., LI, Z., WANG, L., HUANG, Y., WANG, Y. & LI, L. 2019. Rainfall spatial estimations: A review from spatial interpolation to multi-source data merging. *Water,* 11**,** 579.

HUARD, D. & MAILHOT, A. 2006. A Bayesian perspective on input uncertainty in model calibration: Application to hydrological model "abc". *Water Resources Research,* 42.

JEONG, J. & PARK, E. 2017. A shallow water table fluctuation model in response to precipitation with consideration of unsaturated gravitational flow. *Water Resources Research,* 53**,** 3505-3512.

JONES, E., OLIPHANT, T. & PETERSON, P. 2001. SciPy: Open source scientific tools for Python.

JONGJIN, B., JONGMIN, P., DONGRYEOL, R. & MINHA, C. 2016. Geospatial blending to improve spatial mapping of precipitation with high spatial resolution by merging satellite-based and ground-based data. *Hydrological Processes,* 30**,** 2789-2803.

KATIRAIE-BOROUJERDY, P.-S., RAHNAMAY NAEINI, M., AKBARI ASANJAN, A., CHAVOSHIAN, A., HSU, K.-L. & SOROOSHIAN, S. 2020. Bias correction of satellite-based precipitation estimations using quantile mapping approach in different climate regions of Iran. *Remote Sensing,* 12**,** 2102.

KAVETSKI, D., FRANKS, S. W. & KUCZERA, G. 2003. Confronting input uncertainty in environmental modelling. *Calibration of watershed models,* 6**,** 49-68.

KEBEDE, S. 2013. Groundwater occurrence in regions and basins. *Groundwater in Ethiopia.* Springer.

KEBEDE, W., TEFERA, M., HABITAMU, T. & ALEMAYEHU, T. 2014. Impact of land cover change on water quality and stream flow in lake Hawassa watershed of Ethiopia. *Agricultural Sciences,* 2014.

KIDANEMARIAM, S., GOITOM, H. & DESTA, Y. 2021. Coupled application of R and WetSpa models for assessment of climate change impact on streamflow of Werie Catchment, Tigray, Ethiopia. *Journal of Water and Climate Change,* 12**,** 916-936.

KIM, G.-B., CHOI, D.-H. & JEONG, J.-H. 2010. Considerations on the specific yield estimation using the relationship between rainfall and groundwater level variations. *The Journal of Engineering Geology,* 20**,** 61-70.

KRAEMER, H. C. 2014. Kappa coefficient. *Wiley StatsRef: statistics reference online***,** 1-4.

KROES, J., VAN DAM, J., BARTHOLOMEUS, R., GROENENDIJK, P., HEINEN, M., HENDRIKS, R., MULDER, H., SUPIT, I. & VAN WALSUM, P. 2017. SWAP version 4. Wageningen Environmental Research.

KURYLYK, B. L. & MACQUARRIE, K. T. 2013. The uncertainty associated with estimating future groundwater recharge: A summary of recent research and an example from a small unconfined aquifer in a northern humid-continental climate. *Journal of hydrology,* 492**,** 244-253.

LEHNER, F., NADEEM, I. & FORMAYER, H. 2020. An improved statistical bias correction method that also corrects dry climate models. *Hydrology and Earth System Sciences Discussions***,** 1-23.

LEMLEM, A. 2008. *Assessing the impact of land use and land cover change on groundwater recharge using Rs and Gis; a case of Awassa catchement, Southern Ethiopia.* Addis Ababa University.

LI, D., MARSHALL, L., LIANG, Z., SHARMA, A. & ZHOU, Y. 2021. Bayesian LSTM with stochastic variational inference for estimating model uncertainty in process-based hydrological models. *Water Resources Research,* 57**,** e2021WR029772.

LI, J. & HEAP, A. D. 2008. A review of spatial interpolation methods for environmental scientists.

LIN, X. 2020. Promoting the Sustainable Utilization of Groundwater Resources in Ethiopia using the Integrated Groundwater Footprint Index.

LIU, Y. & DE SMEDT, F. 2004. WetSpa extension, a GIS-based hydrologic model for flood prediction and watershed management. *Vrije Universiteit Brussel, Belgium,* 1**,** e108.

LORENTZ, S. A., HUGHES, G. & SCHULZE, R. E. 2003. 11. Techniques for Estimating Groundwater Recharge at Different Scales in Southern Africa. *Groundwater Recharge Estimation in Southern Africa***,** 149.

MA, Q., XIONG, L., XIA, J., XIONG, B., YANG, H. & XU, C.-Y. 2019. A censored shifted mixture distribution mapping method to correct the bias of daily IMERG satellite precipitation estimates. *Remote Sensing,* 11**,** 1345.

MANNA, F., MURRAY, S., ABBEY, D., MARTIN, P., CHERRY, J. & PARKER, B. 2019. Spatial and temporal variability of groundwater recharge in a sandstone aquifer in a semiarid region. *Hydrology & Earth System Sciences,* 23.

MAREI, A., KHAYAT, S., WEISE, S., GHANNAM, S., SBAIH, M. & GEYER, S. 2010. Estimating groundwater recharge using the chloride mass-balance method in the West Bank, Palestine. *Hydrological sciences journal,* 55**,** 780-791.

MARTENS, B., MIRALLES, D. G., LIEVENS, H., VAN DER SCHALIE, R., DE JEU, R. A., FERNÁNDEZ-PRIETO, D., BECK, H. E., DORIGO, W. A. & VERHOEST, N. E. 2017. GLEAM v3: Satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development,* 10**,** 1903-1925.

MCCALLUM, J. L., COOK, P. G. & SIMMONS, C. T. 2015. Limitations of the use of environmental tracers to infer groundwater age. *Groundwater,* 53**,** 56-70.

MCGINNIS, S., NYCHKA, D. & MEARNS, L. O. A new distribution mapping technique for climate model bias correction. Machine Learning and Data Mining Approaches to Climate Science: Proceedings of the 4th International Workshop on Climate Informatics, 2015. Springer, 91-99.

MCKINNEY, W. 2011. pandas: a foundational Python library for data analysis and statistics. *Python for high performance and scientific computing,* 14**,** 1-9.

MCMILLAN, H., JACKSON, B., CLARK, M., KAVETSKI, D. & WOODS, R. 2011. Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models. *Journal of Hydrology,* 400**,** 83-94.

MEYER, P., ROCKHOLD, M. & GEE, G. W. 1997. Uncertainty analyses of infiltration and subsurface flow and transport for SDMP sites. US Nuclear Regulatory Commission (NRC), Washington, DC (United States). Div ….

MOGES, E., DEMISSIE, Y., LARSEN, L. & YASSIN, F. 2021. Sources of Hydrological Model Uncertainties and Advances in Their Analysis. *Water,* 13**,** 28.

MOGES, E., JARED, A., DEMISSIE, Y., YAN, E., MORTUZA, R. & MAHAT, V. Bayesian Augmented L-Moment Approach for Regional Frequency Analysis. World Environmental and Water Resources Congress 2018: Groundwater, Sustainability, and Hydro-Climate/Climate Change, 2018. American Society of Civil Engineers Reston, VA, 165-180.

MOHAMMED, M., BIAZN, B. & BELETE, M. D. 2020. Hydrological Impacts of Climate Change in Tikur Wuha Watershed, Ethiopian Rift Valley Basin. *J Environ Earth Sci,* 10**,** 28-49.

MOHAN, C., WESTERN, A. W., WEI, Y. & SAFT, M. 2018. Predicting groundwater recharge for varying land cover and climate conditions–a global meta-study. *Hydrology and Earth System Sciences,* 22**,** 2689-2703.

MONTANARI, A. & DI BALDASSARRE, G. 2013. The role of model structure to propagate observation uncertainty. *Advances in Water Resources,* 54**,** 242-244.

MORIASI, D. N., ARNOLD, J. G., VAN LIEW, M. W., BINGNER, R. L., HARMEL, R. D. & VEITH, T. L. 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE,* 50**,** 885-900.

MÜLLER-THOMY, H., WALLNER, M. & FÖRSTER, K. 2018. Rainfall disaggregation for hydrological modeling: is there a need for spatial consistence? *Hydrology and Earth System Sciences 22 (2018)*.

MUÑOZ, E., TUME, P. & ORTÍZ, G. 2014. Uncertainty in rainfall input data in a conceptual water balance model: effects on outputs and implications for predictability. *Earth Sciences Research Journal,* 18**,** 69-75.

MURPHY, B. S. PyKrige: development of a kriging toolkit for Python. AGU fall meeting abstracts, 2014. H51K-0753.

MUSTAFA, S. M. T., NOSSENT, J., GHYSELS, G. & HUYSMANS, M. 2020. Integrated Bayesian Multi-model approach to quantify input, parameter and conceptual model structure uncertainty in groundwater modeling. *Environmental Modelling & Software,* 126**,** 104654.

NAVAS, R., SAPRIZA, G., VERVOORT, W. & BAETHGEN, W. On the assimilation of hourly data to scale CHIRPS daily precipitation in Uruguay. Geophysical Research Abstracts, 2019.

NEITSCH, S., ARNOLD, J., KINIRY, J., SRINIVASAN, R. & WILLIAMS, J. 2002. SWAT.

OMONDI, C. K. 2017. *Assessment of bias corrected satellite rainfall products for streamflow simulation: a TOPMODEL application in the Kabompo River Basin, Zambia.* University of Twente.

ORDENS, C. M., WERNER, A. D., POST, V. E., HUTSON, J. L., SIMMONS, C. T. & IRVINE, B. M. 2012. Groundwater recharge to a sedimentary aquifer in the topographically closed Uley South Basin, South Australia. *Hydrogeology Journal,* 20**,** 61-72.

ORKODJO, T. P. 2014. Impact of Land Use/Land Cover Change on Catchment Hydrology (A case Study of Awassa Catchment). *Arba Minch University*.

OWUOR, S. O., BUTTERBACH-BAHL, K., GUZHA, A. C., RUFINO, M. C., PELSTER, D. E., DÍAZ-PINÉS, E. & BREUER, L. 2016. Groundwater recharge rates and surface runoff response to land use and land cover changes in semi-arid environments. *Ecological Processes,* 5**,** 1-21.

PAN, Y., GONG, H., ZHOU, D., LI, X. & NAKAGOSHI, N. 2011. Impact of land use change on groundwater recharge in Guishui River Basin, China. *Chinese Geographical Science,* 21**,** 734-743.

PARK, E. 2012. Delineation of recharge rate from a hybrid water table fluctuation method. *Water Resources Research,* 48.

PIGNONE, F., REBORA, N. & SILVESTRO, F. Modified Conditional Merging technique: a new method to estimate a rainfall field combining remote sensed data and raingauge observations. EGU General Assembly Conference Abstracts, 2015. 3013.

REFSHAARD, J. & STORM, B. 1995. MIKE SHE. *Computer models of watershed hydrology.***,** 809-846.

RIVERBANKCOMPUTING PyQT.

RUTLEDGE, A. T. 2007. Update on the use of the RORA program for recharge estimation. *Groundwater,* 45**,** 374-382.

SCANLON, B. R., HEALY, R. W. & COOK, P. G. 2002. Choosing appropriate techniques for quantifying groundwater recharge. *Hydrogeology journal,* 10**,** 18-39.

SCANLON, B. R., KEESE, K. E., FLINT, A. L., FLINT, L. E., GAYE, C. B., EDMUNDS, W. M. & SIMMERS, I. 2006. Global synthesis of groundwater recharge in semiarid and arid regions. *Hydrological Processes: An International Journal,* 20**,** 3335-3370.

SEIFU, K. 2013. Groundwater in Ethiopia, . *Springer Berlin Heidelberg*.

SIEBERT, S., BURKE, J., FAURES, J.-M., FRENKEN, K., HOOGEVEEN, J., DÖLL, P. & PORTMANN, F. T. 2010. Groundwater use for irrigation—a global inventory. *Hydrology and earth system sciences,* 14**,** 1863-1880.

SIMUNEK, J., VAN GENUCHTEN, M. T. & SEJNA, M. 2005. The HYDRUS-1D software package for simulating the one-dimensional movement of water, heat, and multiple solutes in variably-saturated media. *University of California-Riverside Research Reports,* 3**,** 1-240.

SOLDER, J. E. & BEISNER, K. R. 2020. Critical evaluation of stable isotope mixing end-members for estimating groundwater recharge sources: case study from the South Rim of the Grand Canyon, Arizona, USA. *Hydrogeology Journal,* 28**,** 1575-1591.

SOMARATNE, N. & SMETTEM, K. 2014. Theory of the generalized chloride mass balance method for recharge estimation in groundwater basins characterised by point and diffuse recharge. *Hydrology & Earth System Sciences Discussions,* 11.

SOO, E. Z. X., JAAFAR, W. Z. W., LAI, S. H., OTHMAN, F., ELSHAFIE, A., ISLAM, T., SRIVASTAVA, P. & HADI, H. S. O. 2020. Evaluation of bias-adjusted satellite precipitation estimations for extreme flood events in Langat river basin, Malaysia. *Hydrology Research,* 51**,** 105-126.

STEWART, M., CIMINO, J. & ROSS, M. 2007. Calibration of base flow separation methods with streamflow conductivity. *Groundwater,* 45**,** 17-27.

SWITANEK, M. B., TROCH, P. A., CASTRO, C. L., LEUPRECHT, A., CHANG, H.-I., MUKHERJEE, R. & DEMARIA, E. 2017. Scaled distribution mapping: a bias correction method that preserves raw climate model projected changes. *Hydrology and Earth System Sciences,* 21**,** 2649-2666.

SZILAGYI, J., ZLOTNIK, V. A., GATES, J. B. & JOZSA, J. 2012. Mapping mean annual groundwater recharge in the Nebraska Sand Hills, USA.

TEKLEAB, S., UHLENBROOK, S., MOHAMED, Y., SAVENIJE, H., TEMESGEN, M. & WENNINGER, J. 2011. Water balance modeling of Upper Blue Nile catchments using a top-down approach. *Hydrology and Earth System Sciences,* 15**,** 2179.

TEKLEAB, S., WENNINGER, J. & UHLENBROOK, S. 2014. Characterisation of stable isotopes to identify residence times and runoff components in two meso-scale catchments in the Abay/Upper Blue Nile basin, Ethiopia. *Hydrology and Earth System Sciences,* 18**,** 2415.

TESSEMA, Z. 2004. Hydrogeology of Awassa lake catchment: Isotopic and hydrochemical approach. *Isotope Hydrology and Integrated Water Resources Management***,** 394.

TEUTSCHBEIN, C. & SEIBERT, J. 2012. Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of hydrology,* 456**,** 12-29.

THEMEßL, M. J., GOBIET, A. & HEINRICH, G. 2012. Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. *Climatic Change,* 112**,** 449-468.

TOLERA, M. B. & CHUNG, I.-M. 2021. Integrated Hydrological Analysis of Little Akaki Watershed Using SWAT-MODFLOW, Ethiopia. *Applied Sciences,* 11**,** 6011.

TOLKE, A. A. & AYENEW, T. 2019. The Use of Hydrochemical and Environmental Isotopes in Analysis of Hydrogeological Systems of Abaya-Chamo Lakes Basin. *Journal of Spatial Hydrology,* 15.

TURNADGE, C. & SMERDON, B. D. 2014. A review of methods for modelling environmental tracers in groundwater: Advantages of tracer concentration simulation. *Journal of hydrology,* 519**,** 3674-3689.

VÁZQUEZ-SUÑÉ, E., CARRERA, J., TUBAU, I., SÁNCHEZ-VILA, X. & SOLER, A. 2010. An approach to identify urban groundwater recharge.

VITVAR, T., AGGARWAL, P. & MCDONNELL, J. 2005. A review of isotope applications in catchment hydrology. *Isotopes in the water cycle.* Springer.

VRBA, J. & VAN DER GUN, J. 2004. The world's groundwater resources. *World Water Development Report 2, Contribution to Chapter 4, Report IP 2004-1, System,* 2**,** 1-10.

WALKER, D., PARKIN, G., SCHMITTER, P., GOWING, J., TILAHUN, S. A., HAILE, A. T. & YIMAM, A. Y. 2019. Insights from a multi-method recharge estimation comparison study. *Groundwater,* 57**,** 245-258.

WATSON, A., EILERS, A. & MILLER, J. A. 2020. Recharge estimation using cmb and environmental isotopes in the verlorenvlei estuarine system, south africa and implications for groundwater sustainability in a semi-arid agricultural region. *Water,* 12**,** 1362.

WESTERHOFF, R., WHITE, P. & RAWLINSON, Z. 2018. Incorporation of satellite data and uncertainty in a nationwide groundwater recharge model in New Zealand. *Remote Sensing,* 10**,** 58.

WONDRADE, N., DICK, Ø. B. & TVEITE, H. 2014. GIS based mapping of land cover changes utilizing multi-temporal remotely sensed image data in Lake Hawassa Watershed, Ethiopia. *Environmental monitoring and assessment,* 186**,** 1765-1780.

XIE, Y., COOK, P., SIMMONS, C., PARTINGTON, D., CROSBIE, R. & BATELAAN, O. 2017. Uncertainty of groundwater recharge estimated from a water and energy balance model. *Journal of Hydrology,* 561.

XIE, Y., COOK, P. G., SIMMONS, C. T., PARTINGTON, D., CROSBIE, R. & BATELAAN, O. 2018. Uncertainty of groundwater recharge estimated from a water and energy balance model. *Journal of Hydrology,* 561**,** 1081-1093.

YANG, W., XIAO, C. & LIANG, X. 2019. Analytical sensitivity analysis and uncertainty estimation of baseflow index calculated by a two-component hydrograph separation method with conductivity as a tracer. *Hydrology and Earth System Sciences,* 23**,** 1103-1112.

YE, M., POHLMANN, K. F., CHAPMAN, J. B., POHLL, G. M. & REEVES, D. M. 2010. A model-averaging method for assessing groundwater conceptual model uncertainty. *Groundwater,* 48**,** 716-728.

YENEHUN, A., DESSIE, M., NIGATE, F., BELAY, A. S., AZEZE, M., VAN CAMP, M., TAYE, D. F., KIDANE, D., ADGO, E. & NYSSEN, J. 2022. Spatial and temporal simulation of groundwater recharge and cross-validation with point estimations in volcanic aquifers with variable topography. *Journal of Hydrology: Regional Studies,* 42**,** 101142.

YENEHUN, A., NIGATE, F., BELAY, A. S., DESTA, M. T., VAN CAMP, M. & WALRAEVENS, K. 2020. Groundwater recharge and water table response to changing conditions for aquifers at different physiography: The case of a semi-humid river catchment, northwestern highlands of Ethiopia. *Science of The Total Environment,* 748**,** 142243.

YENENEH, A. 2014. *Characterization of Groundwater-Lake Water Interaction in Lake Hawassa Basin.* Addis Ababa University.

YU, D., YANG, J., SHI, L., ZHANG, Q., HUANG, K., FANG, Y. & ZHA, Y. 2019. On the uncertainty of initial condition and initialization approaches in variably saturated flow modeling. *Hydrology and Earth System Sciences,* 23**,** 2897-2914.

YU, Q., WANG, F., LI, X., YAN, W., LI, Y. & LV, S. 2018. Tracking nitrate sources in the Chaohu Lake, China, using the nitrogen and oxygen isotopic approach. *Environmental Science and Pollution Research,* 25**,** 19518-19529.

ZEMEDAGEGNEHU, L. B. 2020. Groundwater Development and Management Practices in Lake Awassa Catchment, Southern Ethiopia. *Journal of Environment and Earth Science,* 10.

ZHANG, A., SHI, H., LI, T. & FU, X. 2018a. Analysis of the influence of rainfall spatial uncertainty on hydrological simulations using the bootstrap method. *Atmosphere,* 9**,** 71.

ZHANG, L. & DAWES, W. 1998. An integrated energy and water balance model. *CSIRO Land and Water Technical Report.*

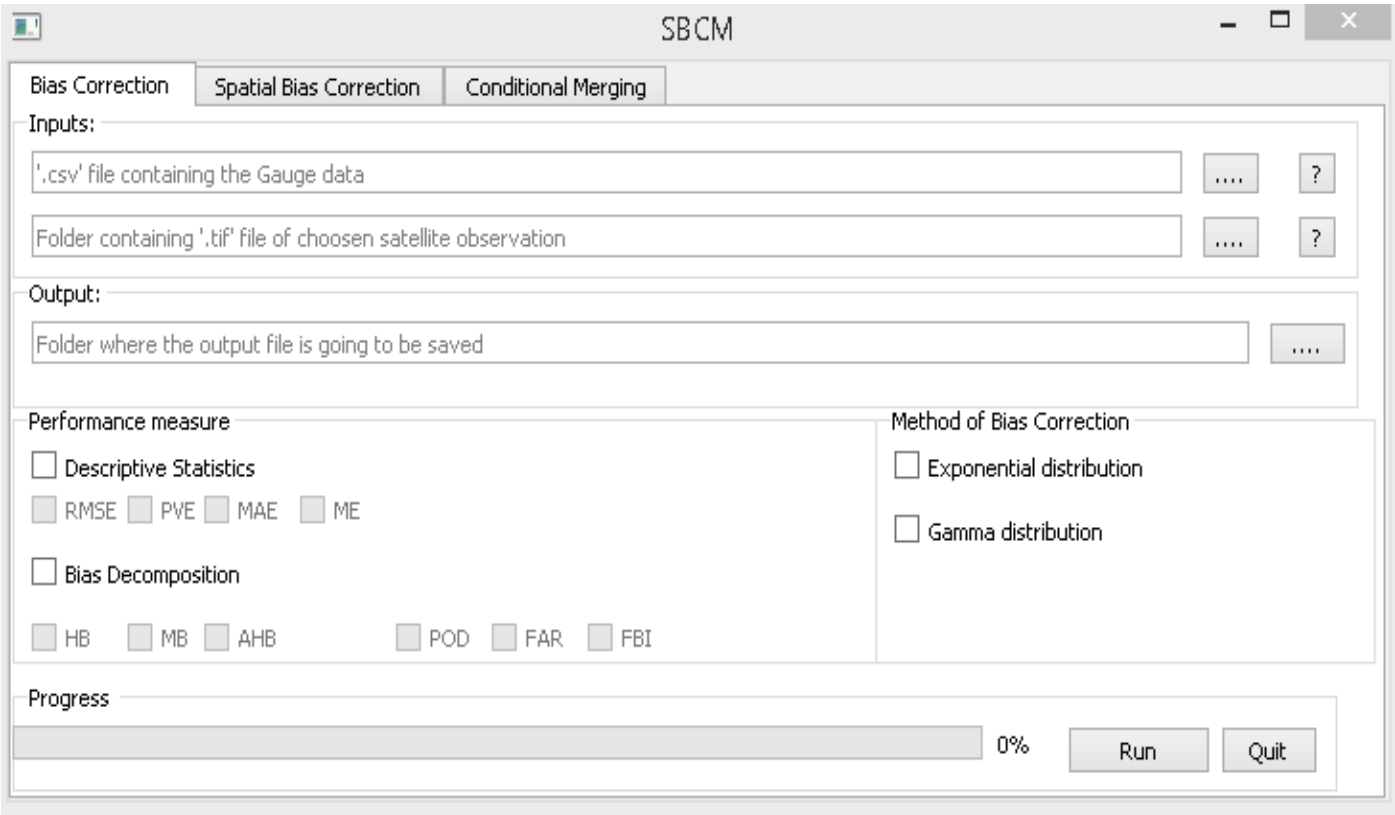
ZHANG, M., ZHI, Y., SHI, J. & WU, L. 2018b. Apportionment and uncertainty analysis of nitrate sources based on the dual isotope approach and a Bayesian isotope mixing model at the watershed scale. *Science of the Total Environment,* 639**,** 1175-1187.

ZHU, R., CROKE, B. F. & JAKEMAN, A. J. 2020. Diffuse groundwater recharge estimation confronting hydrological modelling uncertainty. *Journal of Hydrology,* 584**,** 124642.

Annexes

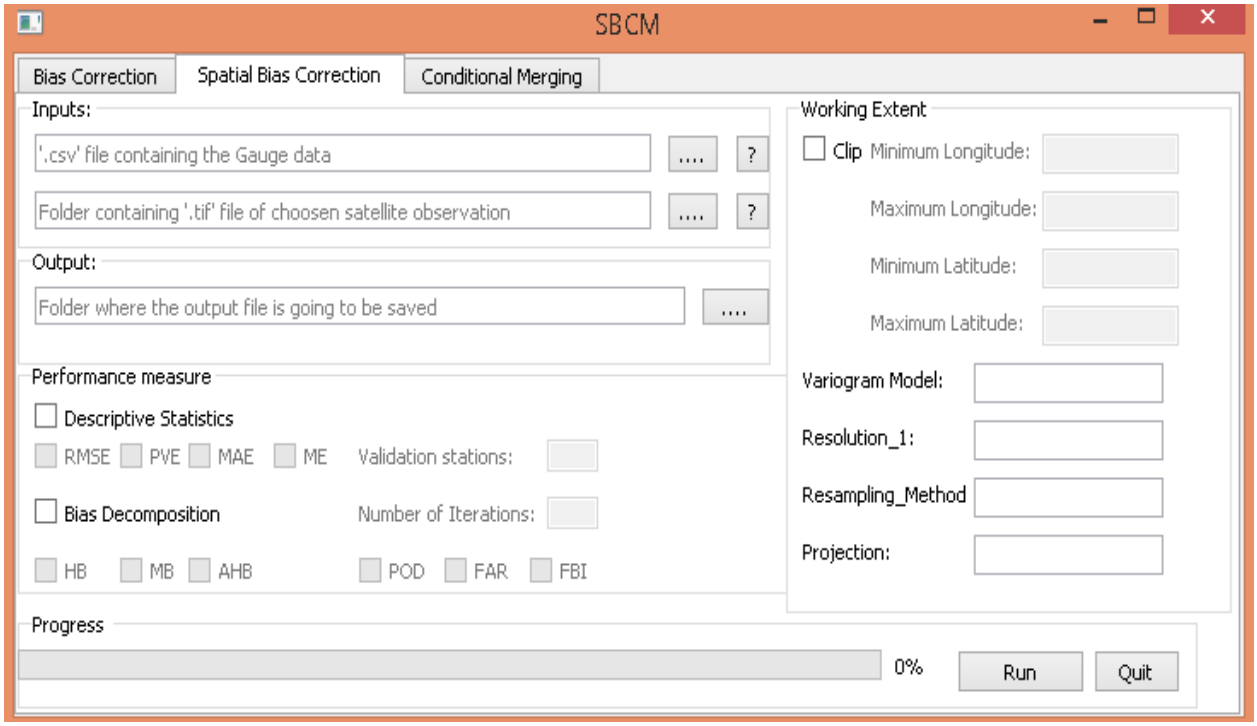
Annex 1: GUI of the developed application for parametric empirical quantile mapping

Inputs	Performance measures	Output
• .csv file of gauge data • .tif file of satellite data	• Descriptive statistics • Bias decompositions	• Bias corrected (daily) .csv file

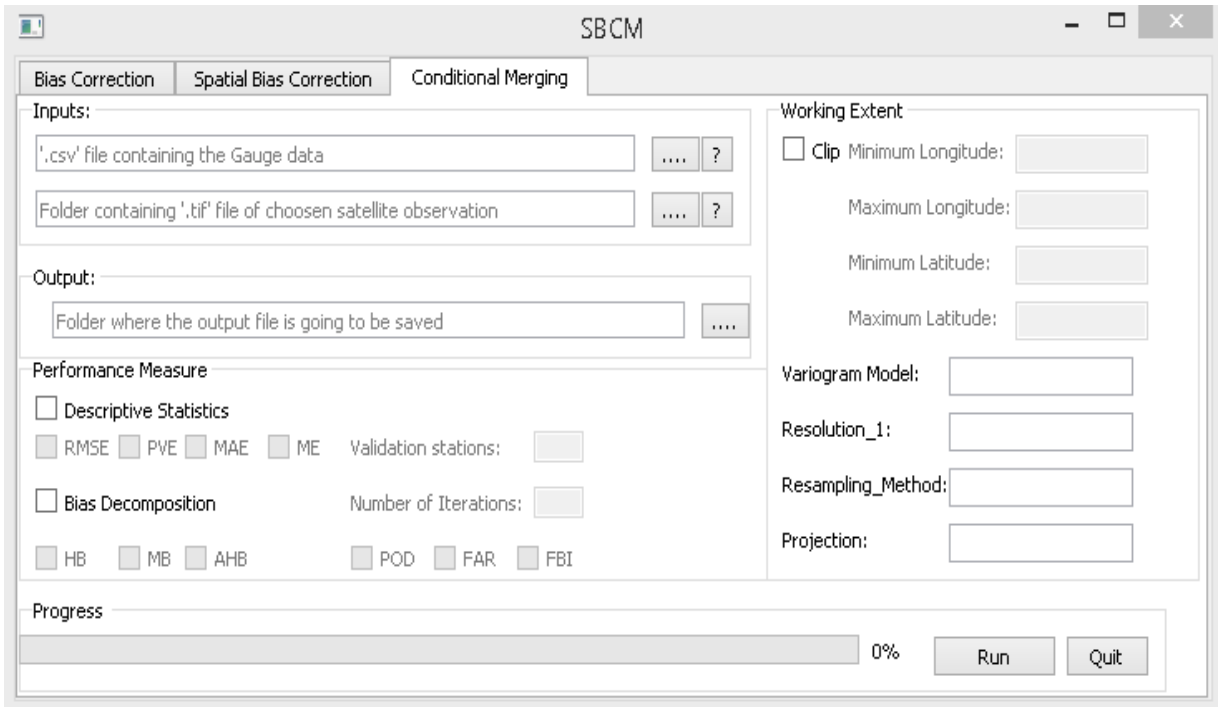


Annex 2: GUI of the developed application for spatial bias corrections

Inputs	Performance measures	Output
• .csv file of gauge data • .tif file of satellite data • .csv file containing parameters at each gauge locations	• Descriptive statistics • Bias decompositions	• Spatially bias corrected (daily) .tif file



Annex 3: GUI of the developed application of conditional merging



Inputs	Performance measures	Output
• .csv file of gauge data • .tif file of satellite data	• Descriptive statistics • Bias decompositions	• Spatially bias corrected and merged (daily) .tif file

Annex 4: Performance evaluation of NN resampling (i.e., to 1 km resolution) techniques for dry and wet periods for CHIRP (V0) and CHIRPS (V2).

Station	Wet period										Dry period									
	RMSE (mm)		PEV (%)		POD		FAR		FBI		RMSE (mm)		PEV (%)		POD		FAR		FBI	
	Vo	V2	Vo	V2	Vo	V2	Vo	V2	Vo	V2	Vo	V2	Vo	V2	Vo	V2	Vo	V2	Vo	V2
BE	8.56	10.6	-21.1	-19.7	0.98	0.45	0.63	0.54	2.64	0.97	4.56	4.81	-89.1	-60.3	0.77	0.29	0.86	0.69	5.31	0.95
HA	7.61	8.97	-3.38	1.81	0.98	0.41	0.5	0.35	1.94	0.63	4.55	4.79	-16.8	9.49	0.86	0.3	0.77	0.46	3.77	0.55
HT	7.33	8.63	-15.3	-9.8	0.98	0.44	0.44	0.28	1.77	0.62	4.3	4.8	-28.5	-5.9	0.85	0.31	0.74	0.39	3.28	0.52
KOF	6.14	8.58	-7.8	-5.3	0.99	0.38	0.27	0.12	1.35	0.43	4.82	5.25	-0.26	14.3	0.9	0.26	0.61	0.19	2.32	0.32
KOK	10.4	12	35.6	38.5	0.99	0.33	0.23	0.16	1.28	0.4	7.42	7.73	50.7	57.3	0.83	0.21	0.54	0.28	1.81	0.3
KU	6.7	8.68	-21	-19.3	0.93	0.36	0.6	0.46	2.32	0.68	4.38	4.83	-56.2	-18.7	0.89	0.28	0.88	0.66	7.67	0.83
LE	7.92	9.82	-25.9	-24.2	0.98	0.46	0.38	0.25	1.57	0.61	4.94	4.96	-13.2	3.02	0.83	0.35	0.69	0.29	2.66	0.5
SHK	6.7	9.07	-42.7	-37.1	0.98	0.4	0.67	0.6	3.01	1	4.58	5.19	-50.9	-17.9	0.81	0.29	0.84	0.59	5.16	0.71
SHA	4.78	7.63	-97.2	-87.2	0.98	0.37	0.48	0.34	1.9	0.56	3.4	4.2	-95	-46.2	0.83	0.25	0.75	0.44	3.36	0.45
TU	7.42	9.8	-20.3	-27.7	0.97	0.44	0.51	0.41	1.98	0.74	5.52	6.09	3.12	11.2	0.82	0.29	0.77	0.55	3.61	0.64
WT	8.38	10.3	-16.1	-13.4	0.97	0.38	0.51	0.47	1.99	0.73	4.56	5.16	-39.9	-21.2	0.86	0.23	0.74	0.59	3.33	0.56
WO	7.82	9.39	1	4.9	0.98	0.41	0.52	0.4	2.04	0.68	4.62	5.17	-4.23	17.7	0.86	0.26	0.78	0.53	3.82	0.55

Annex 5: Performance evaluation of wet period BL resampling (i.e., to 1 km resolution) techniques for CHIRP (V0) and CHIRPS (V2).

Station	Wet Period										Dry Period									
	RMSE (mm)		PEV (%)		POD		FAR		FBI		RMSE (mm)		PEV (%)		POD		FAR		FBI	
	Vo	V2	Vo	V2	Vo	V2	Vo	V2	Vo	V2	Vo	V2	Vo	V2	Vo	V2	Vo	V2	Vo	V2
BE	8.51	10.33	-18.5	-17.4	0.98	0.51	0.63	0.54	2.64	1.11	4.51	4.68	-87.3	-58.8	0.77	0.31	0.86	0.71	5.4	1.07
HA	7.64	8.7	-5.94	-0.19	0.98	0.51	0.5	0.36	1.94	0.79	4.56	4.67	-19.1	6.54	0.86	0.4	0.78	0.45	3.83	0.73
HT	7.27	8.33	-12.7	-8.38	0.98	0.51	0.44	0.29	1.77	0.71	4.26	4.61	-25.4	-4.2	0.84	0.37	0.75	0.4	3.33	0.62
KOF	6.15	8.43	-8.01	-5.29	0.99	0.46	0.27	0.12	1.35	0.52	4.81	5.09	0.49	15.01	0.9	0.35	0.62	0.21	2.36	0.44
KOK	10.44	12.01	35.6	37.48	0.99	0.41	0.23	0.17	1.28	0.5	7.42	7.65	50.76	57.3	0.83	0.26	0.54	0.31	1.81	0.39
KU	6.67	8.09	-20.4	-17	0.93	0.46	0.6	0.47	2.31	0.88	4.37	4.63	-56.1	-10.9	0.89	0.33	0.88	0.66	7.69	0.99
LE	7.91	9.74	-25.2	-23.8	0.98	0.51	0.38	0.26	1.57	0.69	4.94	4.95	-13.4	2.88	0.83	0.38	0.69	0.31	2.67	0.54
SHK	6.68	8.82	-41.6	-35.9	0.98	0.47	0.67	0.6	3.02	1.19	4.58	4.97	-50.6	-18.8	0.81	0.38	0.84	0.59	5.17	0.92
SHA	4.77	7.24	-95.1	-91.7	0.99	0.44	0.49	0.36	1.92	0.69	3.4	3.94	-95.7	-54.5	0.88	0.31	0.76	0.43	3.74	0.55
TU	7.49	9.52	-24	-23.7	0.97	0.49	0.51	0.42	1.98	0.84	5.54	5.95	0.54	14.08	0.82	0.34	0.77	0.56	3.64	0.78
WT	8.32	10.23	-16.8	-13.8	0.97	0.47	0.52	0.46	2.01	0.87	4.42	5.06	-40	-21.5	0.9	0.3	0.75	0.59	3.53	0.74
WO	7.8	9.14	1.56	5.1	0.99	0.5	0.52	0.41	2.04	0.84	4.6	5.01	-2.72	18.54	0.86	0.31	0.78	0.57	3.87	0.73

Annex 6: Performance evaluation of separate BL resampling (i.e., to 1 km resolution) techniques for wet and dry periods using bias decomposition.

Station	Wet period				Dry Period			
	HB(mm)		MB(mm)		HB (mm)		MB (mm)	
	Vo	V2	Vo	V2	Vo	V2	Vo	V2
BE	-4414.6	-214.3	-206.8	-4197.6	-428	151	-250.9	-758.6
HA	-3225.7	645.9	-90.8	-3513.5	-668.9	149.8	-228.4	-884.4
HT	-2258.1	1695.1	-61.3	-3381.9	-532.5	455.9	-216.3	-981
KOF	-1132.2	3444.7	-14.8	-3973.9	-829.1	726.7	-178	-1357.6
KOK	-7337.3	1203.3	-88.6	-8578.8	-2797.2	-187	-504.4	-3008.5
KU	-2557.6	510	-429.3	-3078.2	-715.6	-131.8	-105.9	-621.5
LE	-911.2	3261.9	-128.4	-3746.5	-506	600.1	-270.2	-1136.4
SHK	-3119	119	-110.8	-3438.6	-652.7	239	-304.6	-939.1
SHA	589.2	3322.6	-34.9	-2152.1	-118.8	585	-82.5	-586.4
TU	-2420.8	1557.4	-190.4	-4040.5	-1127.6	51.6	-395.8	-1546.6
WT	-3683.4	911	-161.9	-4681.9	-708.7	398.7	-162.4	-1264.6
WO	-4357.8	249	-124	-4345	-1094.8	36.3	-260.8	-1365.6

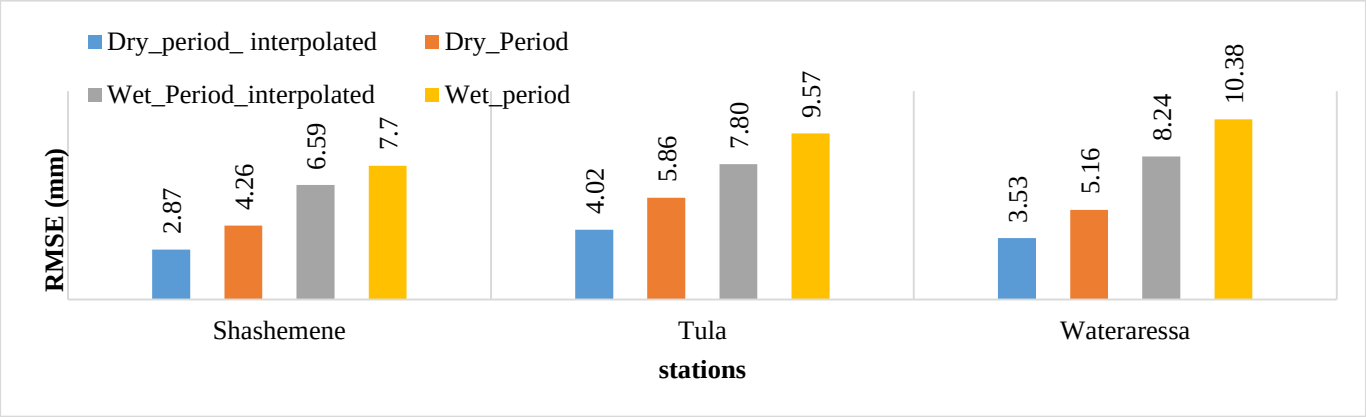
Annex 7: Performance evaluation of NN resampling (i.e., to 1 km resolution) techniques for wet and dry periods separately using bias decomposition.

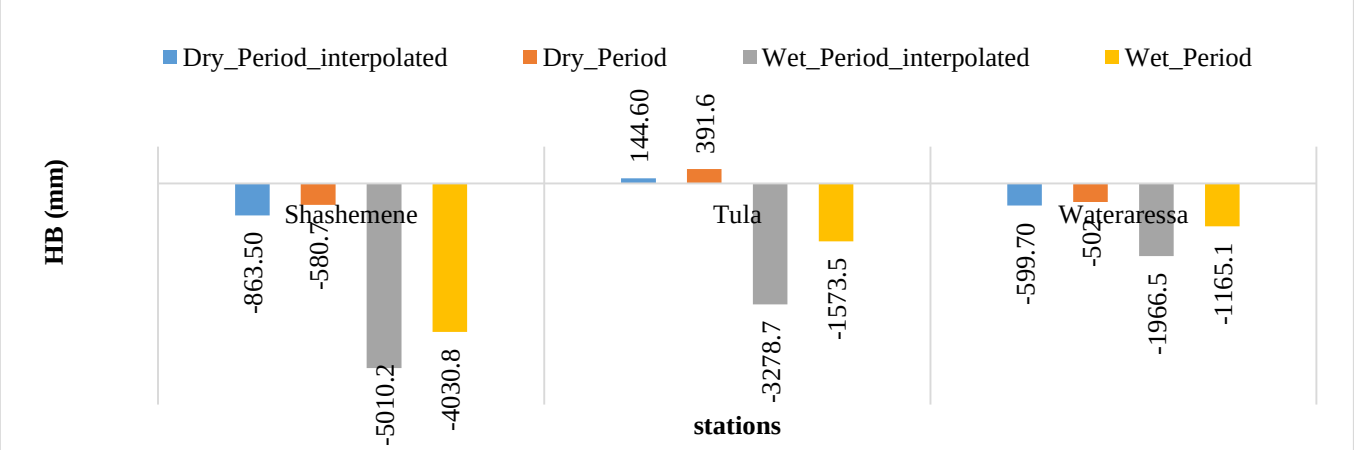
Station	Wet Period				Dry Period			
	HB(mm)		MB(mm)		HB (mm)		MB (mm)	
	Vo	V2	Vo	V2	Vo	V2	Vo	V2
BE	-4311.8	-1837.5	-204.8	-4826.1	-412.8	-843.6	-250.9	-788.1
HA	-3352.1	159.8	-90.8	-4309.9	-692.2	181.2	-228.4	-1077.6
HT	-2122.9	-837.3	-61.3	-4022.7	-505.7	-112.2	-216.1	-1113
KOF	-1147.8	-494.6	-14.8	-4756.9	-820.7	392.4	-173.1	-1673.1
KOK	-7337.3	6027.6	-88.6	-9896.4	-2797.2	2715.1	-504.4	-3299.9
KU	-2541.7	-1330.8	-429.3	-3922.2	-715.1	-235.5	-105.9	-712.6
LE	-865.7	-2253.1	-129.8	-4137.3	-510.3	70.3	-266.8	-1203.8
SHK	-3053.9	-2658.8	-134.4	-3948	-650.7	-303.5	-304.6	-1094.2
SHA	682.2	-3927.4	-51.5	-2495.1	-81.8	-531.9	-122.6	-709.8
TU	-2595.2	-2437.5	-192.9	-4496.7	-1166.2	291.1	-395.8	-1694.8
WT	-3675.2	-1249.6	-198.5	-5404.1	-603.7	-423.2	-259	-1419
WO	-4328.3	469.1	-125	-5351.8	-1077	394.9	-260.8	-1465.7

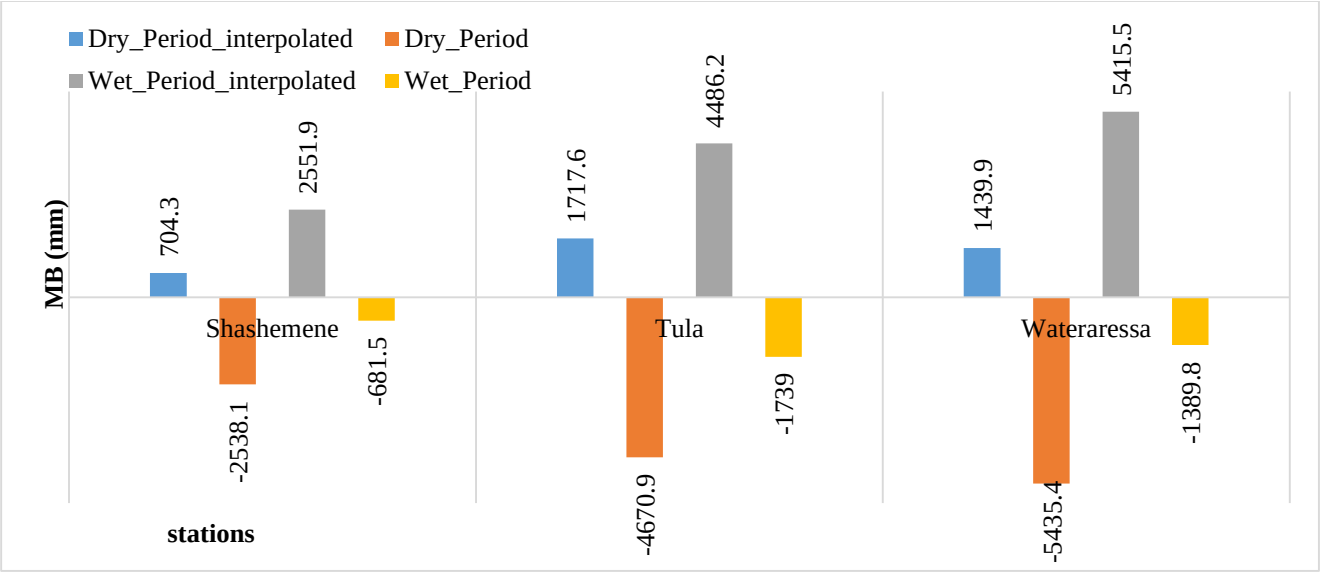
Annex 8: Bias decomposition and descriptive statistics test for bias-corrected CHIRPS after resampling (NN) and at available collocated ground stations

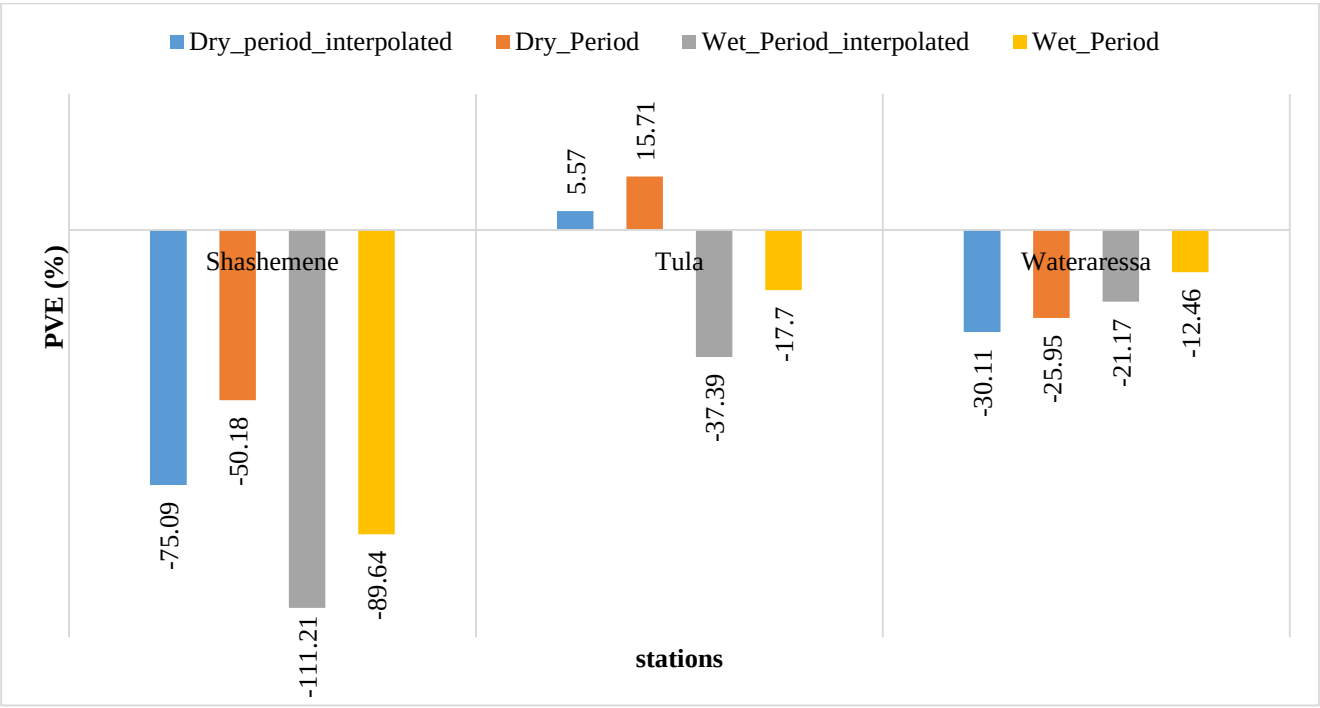
Station	RMSE (mm)		PVE (%)		HB (mm)		MB(mm)	
	Dry period	Wet period	Dry period	Wet period	Dry period	Wet period	Dry period	Wet period
BE	4.9	10.58	-47.67	-21.43	-701.9	-1979.2	-4807.3	-806.9
HA	4.79	8.97	10.58	1.58	201.3	139.7	-4313.6	-1073.9
HT	4.8	8.63	-6.07	-9.77	-113.3	-836.2	-4030	-1105.7
KOF	5.25	8.58	12.4	-4.7	332.7	-434.9	-4780.6	-1649.4
KOK	7.66	12.07	57.85	38.39	2734.2	6008.5	-9934.9	-3261.4
KU	4.83	8.68	-15.47	-19.92	-197	-1369.3	-3904.9	-729.9
LE	4.86	9.86	4.83	-24.73	112.2	-2295	-4156.7	-1184.4
SHK	5.1	9.1	-19.95	-36.51	-326	-2636.3	-3977.7	-1064.5
SHA	4.26	7.7	-50.18	-89.64	-580.7	-4030.8	-2538.1	-681.5
TU	5.86	9.57	15.71	-17.7	391.6	-1573.5	-4670.9	-1739
WT	5.16	10.38	-25.95	-12.46	-502	-1165.1	-5435.4	-1389.8
WO	5.18	9.39	16.46	5.27	358.4	505.6	-5381.8	-1435.7

Annex 9: Performance test for CHIRPS using interpolated distribution parameters






Dry_Period_interpolated
Dry_Period
Wet_Period_interpolated
Wet_Period
MB (mm)
704.3
-2538.1
2551.9
-681.5
1717.6
-4670.9
4486.2
-1739
1439.9
-5435.4
5415.5
-1389.8
Shashemene
Tula
Wateraressa
stations


Dry_period_interpolated
Dry_Period
Wet_Period_interpolated
Wet_Period
PVE (%)
-75.09
-50.18
-111.21
-89.64
5.57
15.71
-37.39
-17.7
-30.11
-25.95
-21.17
-12.46
Shashemene
Tula
Wateraressa
stations

Annex 10: Performance of the exponential distribution for station bias correction results

Station	RMSE (mm)		PVE (%)		HB (mm)		MB (mm)		POD		FAR		FBI	
	Dry	Wet	Dry	Wet	Dry	Wet	Dry	Wet	Dry	Wet	Dry	Wet	Dry	Wet
BE	5.6	7.63	-42.76	3.03	-598	281.7	-250.9	-204.8	0.77	0.98	0.4	0	1.29	0.98
HA	5.58	7.13	-47.69	3.16	-910.2	279.5	-228.4	-90.8	0.86	0.98	0.4	0	1.43	0.98
HT	4.72	7.09	-39.39	0.91	-740.7	77.9	-216.1	-61.3	0.85	0.98	0.4	0	1.42	0.98
KOF	5.55	5.94	-30.33	0.14	-829.2	12.6	-173.1	-14.8	0.9	0.99	0.32	0	1.32	0.99
KOK	9.03	10.91	-18.53	0.83	-876.6	130.3	-504.4	-88.6	0.83	0.99	0.31	0	1.2	0.99
KU	4.76	6.21	-61.87	0.14	-778.2	9.6	-135.5	-76.8	0.85	0.99	0.48	0	1.64	0.99
LE	5.33	7.63	-26.66	1.14	-619.9	105.8	-266.8	-129.8	0.83	0.98	0.35	0	1.28	0.98
SHK	5.49	5.87	-61.55	0.13	-1038.9	9.4	-241.7	-82.4	0.86	0.99	0.45	0	1.57	0.99
SHA	3.06	3.5	-40.96	7.02	-471	316	-149.3	-276.9	0.83	0.94	0.39	0	1.37	0.94
TU	6.75	7.29	-33.89	0.98	-879.4	86.5	-392.7	-192.5	0.82	0.97	0.41	0	1.39	0.97
WT	5.36	7.68	-33.5	2.87	-667.3	266.6	-259	-198.5	0.86	0.97	0.37	0	1.36	0.97
WO	6.34	7.36	-56.04	1.92	-1249.9	183.3	-273.4	-125	0.85	0.98	0.45	0	1.54	0.98

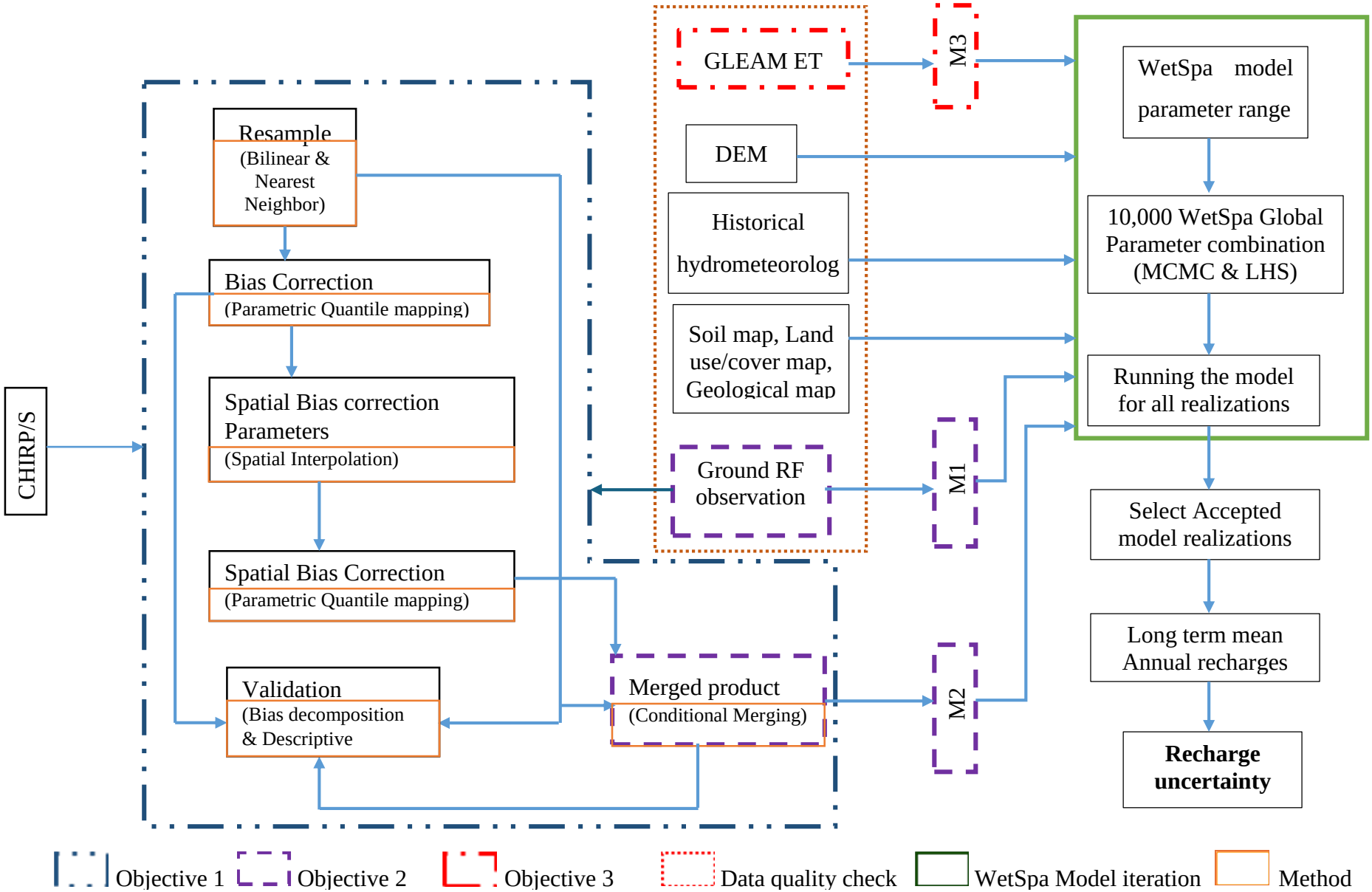
Annex 11: Performance of the gamma distribution for station bias correction results

Station	RMSE (mm)		PVE (%)		HB (mm)		MB (mm)		POD		FAR		FBI	
	Dry	Wet	Dry	Wet	Dry	Wet	Dry	Wet	Dry	Wet	Dry	Wet	Dry	Wet
BE	3.78	7.81	0.01	11.94	0.1	1111.2	-484.1	-924.9	0.56	0.89	0.38	0	0.9	0.89
HA	5.46	8.45	2.48	11.92	47.4	1054.7	-636.2	-658.7	0.63	0.89	0.37	0	1.05	0.89
HT	4.55	8.09	1.72	11.73	32.4	1002.3	-554.5	-635.7	0.62	0.88	0.37	0	0.99	0.88
KOF	5.41	6.8	10.92	11.96	298.4	1100.4	-644.4	-728.1	0.67	0.89	0.25	0	0.89	0.89
KOK	8.72	12.47	30.75	13.85	1454.8	2167.5	-1716.2	-1613.3	0.54	0.88	0.25	0	0.72	0.88
KU	4.84	6.68	-30.65	10.34	-385.5	712.4	-264.3	-636.3	0.71	0.9	0.45	0	1.29	0.9
LE	5.01	8.75	10.12	10.98	235.3	1018.2	-751.8	-894	0.62	0.87	0.3	0	0.89	0.87
SHK	4.52	5.47	-16.45	12.99	-277.7	931	-582.8	-787.6	0.6	0.87	0.45	0	1.07	0.87
SHA	2.45	3.8	14.42	17.54	165.8	789.8	-462.2	-659.5	0.54	0.84	0.35	0	0.82	0.84
TU	5.83	7.42	7.48	10.56	194.2	928.3	-1032.6	-819.4	0.57	0.89	0.38	0	0.92	0.89
WT	3.62	8.4	7.33	14.27	146	1326.5	-805.2	-1156	0.57	0.86	0.35	0	0.87	0.86
WO	4.81	8.11	-1.95	12.37	-43.5	1180.1	-884.1	-1010.7	0.58	0.88	0.41	0	0.97	0.88

Annex 12: Performance of conditionally merged CHIRP SREs after bias correction

St.	Dry Period							Wet Period						
	RMSE (mm)	PVE (%)	HB (mm)	MB (mm)	POD	FAR	FBI	RMSE (mm)	PVE (%)	HB (mm)	MB (mm)	POD	FAR	FBI
BE	3.52	-53.36	-746.3	-47.9	0.93	0.74	3.61	8.14	-13.04	-1213.4	-116.4	0.99	0.63	2.62
HA	3.28	0.68	12.7	-111.1	0.86	0.54	1.86	5.72	17.42	1488.1	-414	0.89	0.37	1.42
HT	3.66	-16.3	-311.1	-66.8	0.94	0.6	2.34	6.26	4.79	-325	-32.2	0.98	0.47	1.86
KOF	4.01	-10.17	-278	-36.6	0.91	0.48	1.75	5.86	-8.75	-805.6	-195.9	0.95	0.25	1.28
KOK	6.91	45.89	2171.2	-671.3	0.78	0.36	1.23	10.04	34.33	5371.2	-786.2	0.92	0.22	1.19
KU	3.84	-52.92	-665.6	-70.8	0.87	0.79	4.35	6.67	-17.77	-1224.3	-717.8	0.8	0.57	1.88
LE	3.94	-17.64	-410.3	-17.4	0.95	0.53	2.05	7.05	-19.8	-1836.9	-26	0.98	0.37	1.57
SHK	3.84	-17.64	-410.3	-17.4	0.87	0.53	2.05	6.94	-17.77	-1224.3	-1235.7	0.8	0.37	1.57
SHA	2.55	-81.45	-936.6	-38.5	0.9	0.55	2	4.8	-72.97	-3286.7	-552.6	0.81	0.46	1.52
TU	4.52	9.78	253.8	-111.9	0.9	0.61	2.29	6.93	-15.77	-1386.5	-81.7	0.98	0.5	1.96
WT	3.05	-6.51	-129.7	-152.8	0.81	0.56	1.84	7.85	-5.84	-543.2	-226	0.94	0.51	1.92
WO	3.85	15.7	350.2	-333.6	0.79	0.63	2.14	7.01	16.29	1553.4	-379.4	0.92	0.48	1.78

Annex 13: General methodology flow chart



RESEARCH OUTPUTS

JOURNAL PUBLICATIONS (In SCI- Indexed Journals)

1. **BEYENE, T. D.**, ZIMALE, F. A. & GEBREKRISTOS, S. T. 2024a. A review on sources of uncertainties for groundwater recharge estimates: insight into data scarce tropical, arid, and semiarid regions. Hydrology Research, 55, 51-66. https://doi.org/10.2166/nh.2023.221

2. **BEYENE, T. D.**, ZIMALE, F. A., GEBREKRISTOS, S. T. & NEDAW, D. 2023. Evaluation of a multistaged bias correction approach on CHIRP and CHIRPS rainfall product: a case study of the Lake Hawassa watershed. Journal of Water and Climate Change, 14, 1847-1867. https://doi.org/10.2166/wcc.2023.457

3. **BEYENE, T. D.**, ZIMALE, F. A., GEBREKRISTOS, S. T. & NEDAW, D. 2024b. Assessment of the impact of rainfall uncertainties on the groundwater recharge estimations of the Tikur-Wuha watershed, rift valley lakes basin, Ethiopia. Heliyon, 10. https://doi.org/10.1016/j.heliyon.2024.e24311