



**Addis Ababa University, Addis Ababa Institute of
Technology (AAiT)**

**Evaluation of Global Water Resources Data Products for Local
Water Management Application in the Upper Blue Nile Basin in
Ethiopia**

By

Haileyesus Belay Lakew

Supervisor: Semu Ayalew Moges (PhD, Assoc. Prof.)

Dereje Hailu Asfaw (PhD, Assoc. Prof.)

PhD Dissertation submitted to Addis Ababa University, Addis Ababa Institute of Technology,
School of Civil and Environmental Engineering in Partial Fulfillment of the Doctor of Philosophy
(PhD) in Civil (Hydraulic) Engineering.

Addis Ababa University

Addis Ababa, Ethiopia

2018

Addis Ababa University (AAU), Addis Ababa Institute of Technology (AAiT)

Doctoral Dissertation Approval Sheet

Evaluation of Global Water Resources Data Products for Local Water Resources Management Application in the Upper Blue Nile Basin in Ethiopia

By:

Haileyesus Belay Lakew

Date

APPROVED BY

BOARD OF EXAMINERS:

Advisor: - Semu Ayalew Moges (Assoc. Prof.)

Date

Co-Advisor: - Dereje Hailu Asfaw (Assoc. Prof.)

Date

External Examiner: - Paul Block (Assis. Prof.)

Date

Internal Examiner 1: - Seifu Tilahun (Assoc. Prof.)

Date

Internal Examiner 2: - Yilma Sileshi (Assoc. Prof.)

Date

Chairperson: - Agizew Nigussie (Assis..Prof.)

Date

Abstract

Under the condition of water scarcity and competing water uses, improved knowledge of basin wide hydrology is pivotal for various water resources management and applications. Availability and quality of hydro-meteorological data sets determine understanding of the hydrology and sustainable planning and management of the basin resources. The quantity and quality of these data vary from region to region, and observations are sparse and irregularly distributed and declining in many countries particularly in Africa. To overcome these challenges and generate a consistent set of hydro-meteorological data products for global climatic and environmental understanding, alternative sources of data are emerging from different global research institutions. The increasing availability of global observation datasets, both from in-situ and remote sensors, along with advancements in earth system models and data assimilation algorithms, have led to the generation of a number of water resources reanalysis products that are available at global scale with high spatial and temporal resolutions. These global products hold great potential for water resource application at local scales in different parts of the world.

The aim of this study is to evaluate these global hydro-meteorological data products and improve the uncertainty for local use in local water resources application in upper Blue Nile basin, Ethiopia. The study has two components. The first component consists of evaluation different global precipitation products for runoff estimation at local level. The second component consists of evaluation of different global available runoff products for local water resources management application. In the first case, a fully distributed hydrological model Coupled Routing and Excess Storage (CREST) is used at different spatio-temporal scales. Two satellite and three reanalysis global gridded precipitation products with resolution of 0.5° and 0.25° are used.

In the second case, Water Resources Reanalysis (WRR) river runoff products, WRR1 (seven products) available at 0.5° and WRR2 (four products) available at 0.25° are evaluated. Based on uncertainty analysis of these products, different dynamical bias correction schemes are implemented at each grid to improve the quality of the river runoff data for application at local scale.

Overall, the results show, the global precipitation and runoff products improve tremendously and can be used as alternative sources of data for various water resources applications with limited local fine tuning and bias correction. For the upper Blue Nile basin, the study indicates evaluation and adaptation of global reanalysis datasets is needed before application of the data for local water resources planning, design and management decisions. Using dynamical bias correction implemented at each grid (0.25°), this study develops 30 years (1980-2010) consistent set of daily streamflow data which can be used for various water management applications at local scale for both gauged and ungauged watersheds in the basin.

ACKNOWLEDGEMENTS

First and foremost, thanks to the Almighty God for granting me His limitless care, love and blessings all along the way.

I would like to express my gratitude toward those who have helped and supported me in a variety of different manners during my time as a doctoral student at the Addis Ababa Institute of Technology.

I am grateful and express my deepest gratitude to my advisors Dr. Semu Ayalew Moges and Dr. Ing. Dereje Hailu for their unreserved assistance, constructive and timely comments at all stages of my work and also for supplying me relevant materials to carry out the research. I should strongly appreciate Dr. Semu for his patience full guidance in a lot of discussions we made on various problems I faced during the course and research of the work.

I would like to acknowledge EU funded earth2Observe project (ENVE.2013.6.3-3) without their full financial and data support my PhD wouldn't have been realized. I would like to take this opportunity to extend my thanks to the School of Civil and Environmental Engineering, AAiT, AAU for hosting and facilitating my study and Addis Ababa University for providing travel and partial living cost to my short scientific visit to University of Connecticut (UCONN).

I would like to acknowledge Prof. Emmanouil N. Anagnostou for arranging a short scientific visit to the University of Connecticut (UCONN), USA. I would like to extend my appreciation to Prof. Efthymios I. Nikolopoulos, for supporting me during my stay at UCONN. I would also like to express my thanks to Prof. George Kallos for hosting me in the University of Athens, Greece for general training in distributed hydrological model (CREST), processing of global precipitation products and Matlab programming language during my early stage of my PhD.

I would like to acknowledge Ethiopian Ministry of Water, Irrigation and Electricity and the National Meteorological Agency for providing me the required data and information to fulfill this research work.

Lastly, I would wish to convey my warm feeling of appreciation and thanks to my wife and families who are constantly encouraging and providing me the necessary care and love.

Table of Contents

Abstract.....	ii
Table of Contents	vi
List of Figures.....	x
List of Tables	xii
Chapter One	1
1. General Introduction and Background.....	1
1.1. Background and Motivation	1
1.2. Research Questions	6
1.3. Research Objectives	7
1.4. Thesis Structure.....	7
Chapter Two.....	10
2. Data and Study Methodology	10
2.1. Description of Study Area	10
2.2. In Situ Hydro-Meteorological Data	11
2.3. Global Precipitation Products.....	18
2.4. Global River Runoff Products	20
2.4.1. Spatial Issue of Global Runoff Products.....	25
2.4.2. Temporal Issue of Global Runoff Products.....	29
2.5. Subject for Revision	30

2.6.	Global Potential Evapotranspiration	33
2.7.	CREST Model.....	34
2.7.1.	Overview of CREST	34
2.7.2.	New in CREST V2.1	36
2.8.	Parameters	38
2.9.	Performance in Discharge Simulation.....	40
Chapter Three		43
3.	Evaluation of Satellite and Reanalysis Precipitation Products at Small scale: A Case Study of Gilgel Abbay.....	43
3.1.	Introduction	43
3.2.	Data and Methods	45
3.2.1.	Study Area	45
3.2.2.	Data Sets	46
3.2.3.	Parameters and Calibration.....	47
3.3.	Result and Discussion	50
3.4.	Conclusive Remarks.....	59
Chapter Four.....		62
4.	Multi-Scale Evaluation of Globally Available Satellite and Reanalysis Precipitation Products in the Upper Blue Nile Basin (Abbay), Ethiopia.....	62
4.1.	Introduction	62

4.2. Data and Methods	65
4.2.1. Study Area	65
4.2.2. Data Sets	68
4.2.3. Parameters and Calibration	68
4.3. Result and Discussion	70
4.3.1. Error of precipitation propagation to the simulated flow	84
4.4. Conclusive Remarks	86
Chapter Five	89
5. Evaluation of Global Water Resources Reanalysis Runoff Products for Local Water Resources Applications: A case of Upper Blue Nile Basin in Ethiopia.	89
5.1. Introduction	89
5.2. Data and Methods	92
5.2.1. Study Area	92
5.2.2. Data Sets	92
5.3. Result and Discussion	94
5.4. Conclusive Remarks	114
Chapter Six	116
6. Dynamical Bias Correction Procedure to Improve the WRR Gridded Daily Streamflow Data for Local Application in the Upper Blue Nile Basin	116
6.1. Introduction	116

6.2. Data and methods	117
6.2.1. Study Area	117
6.2.2. Data Sets	118
6.2.3. Scheme for Bias Correction	120
6.3. Result and Discussion	123
6.3.1. Evaluation of Bias Correction Schemes	124
6.3.2. Evaluation of the Selected Bias Correction Scheme and Runoff	126
6.4. Conclusive Remarks	133
Chapter Seven	135
7. Summary, Conclusions and Recommendations	135
7.1. Summary and Conclusions	135
7.2. Recommendations	141
References	143
Appendices	152

List of Figures

Figure 1. Location of sub-basins in the upper Blue Nile (Abbay) basin.	11
Figure 2. The daily mean Hydrograph of great and little Anger in the upper Blue Nile basin. ...	13
Figure 3. The daily mean hydrograph of Jema watershed in the upper Blue Nile basin.	14
Figure 4. Dry season monthly mean streamflow of Gilgel Abbay watershed in the upper Blue Nile basin.	15
Figure 5. Monthly streamflow of Kessie for the updated and former rating curves.	16
Figure 6. Monthly streamflow hydrograph of Eldiem for Ethiopian and Sudanese stations.	17
Figure 7. Stream network of the upper Blue Nile (Abbay) basin with different reanalysis river runoff products (WRR1, 0.5°).	29
Figure 8. High resolution (0.25°) reanalysis river runoff product of UniK with the existing network of the upper Blue Nile basin.	33
Figure 9. Potential Evapotranspiration derived from Penman-Monteith of the Upper Blue Nile (Abbay) basin.	34
Figure 10. Core components of the CREST model	35
Figure 11. Routing concept (a) Linear reservoir routing (LRR) method and (b) Fully distributed reservoir (DLRR) method	36
Figure 12. Gilgel Abbay watershed with locations of five nearby rain gauge stations of the study region, and one stream gauge station at the outlet of the watershed with 0.25° grid resolution. .	46
Figure 13. Hydrograph of simulated flow of global precipitation products with observed flow of Gilgel Abbay watershed for the Independent calibration mode (a,b) CMORPH, (c,d) TRMM, (e,f) ECMWF.	55

Figure 14. Different precipitation products comparison of (a,b) Average annual rainfall, (c,d) Average annual flow for Simulation mode, (e,f)Average annual flow for Independent calibration mode, (g,h) Average annual flow for RainFact calibration mode.	57
Figure 15. Quantile-Quantile plot of Gilgel Abbay watershed for (a) Independent calibration, (b) In-situ gauge parameter simulation, (c) RainFact calibration.....	59
Figure 16. Location of the three nested watersheds and rain gauge stations distribution in the upper Blue Nile basin.....	67
Figure 17. Mean annual rainfall of (a) Gilgel Abbay (b) Kessie and (c) Eldiem case studies.	72
Figure 18. Daily Quantile-Quantile plots of Gilgel Abbay (a) simulation mode (b) RainFact calibration mode, Kessie (c) simulation mode (d) RainFact calibration mode, Eldiem (e) simulation mode (f) RainFact calibration mode	83
Figure 19. Location of the three case studies in upper Blue Nile basin.....	92
Figure 20. Taylor diagram of (a) Daily, (b) Monthly and (c) Seasonal runoff of Gilgel Abbay watershed.	98
Figure 21. Taylor diagram of (a) Daily, (b) Monthly and (c) Seasonal runoff of Kessie.....	101
Figure 22. Taylor diagram of (a) Daily, (b) Monthly and (c) Seasonal runoff of Eldiem.....	104
Figure 23. WRR1 QQ-plot of Gilgel Abbay watershed (a) Daily (b) Monthly (c) Seasonal, Kessie station (d) Daily (e) Monthly (f) Seasonal, Eldiem Station (g) Daily (h) Monthly (i) Seasonal runoff.....	107
Figure 24. WRR2 QQ-plot of Gilgel Abbay watershed (a) Daily (b) Monthly (c) Seasonal, Kessie station (d) Daily (e) Monthly (f) Seasonal, Eldiem Station (g) Daily (h) Monthly (i) Seasonal runoff.....	108

Figure 25. Critical Success Index (CSI) of the 50th-percentile seasonal flows of Gilgel Abbay, Kessie and Eldiem basin scales.....	110
Figure 26. Critical Success Index (CSI) of the 80th-percentile monthly flow of Gilgel Abbay, Kessie and Eldiem basin scales.....	111
Figure 27. Location of streamflow stations in the upper Blue Nile basin.	118
Figure 28. QQ plots of the corrected and uncorrected UniK products for different stations.	132

List of Tables

Table 1. Parameter values of CREST model.	39
Table 2. Estimated parameter values of independent calibrations for different precipitation products.....	48
Table 3. Different calibration and simulation modes results of global precipitation products and gauged rainfall.	52
Table 4. Estimated parameter values of independent calibration for different case studies.	69
Table 5. Daily performance of various precipitation products of the Gilgel Abbay watershed. ..	75
Table 6. Daily performance of various precipitation products of Kessie.	77
Table 7. daily performance of various precipitation products of Eldiem.	79
Table 8. Error (bias %) of satellite and reanalysis precipitation products propagates to the simulated flow from 2000 to 2012.	85
Table 9. NSCE performance of WRR products and simulated runoff of Gilgel Abbay watershed (small-scale 1,656 km ²).	95
Table 10. NSCE performance of WRR products and simulated flows of Kessie (medium-scale 65,784 km ²).....	99

Table 11. NSCE performance of WRR products and simulated flows of Eldiem (large-scale 199,812 km ²).....	102
Table 12. Performance of global WRR2 (UniK) product for different case studies.	112
Table 13. Different schemes of bias correction for the three-different sized watersheds from 2000 to 2010.	125
Table 14. Performance of the new set of bias corrected UniK (BF _{TSV}) river discharge from 1980 to 2010.	128

Abbreviations

AAiT	Addis Ababa Institute of Technology
ACPC	African Climate Policy center
ASCE	American Society of Civil Engineering
AWRASL	Australian Water Resources Assessment System Landscape
BCM	Billion Cubic Meter
BF	Bias Factor
CC	Correlation Coefficient
CNRS	Centre National de la Recherche Scientifique
CREST	Coupled Routing and Excess Storage
CSI	Critical Success Index
CSIRO	Commonwealth Scientific and Industrial Research Organization
DEM	Digital Elevation Model
DWD	Deutscher Wetter Dienst
ECMWF	European Central for Medium- Range Weather Forecasting
EO	Earth Observation
ERA-Interim	ERA-Interim
FAO	Food and Agriculture Organization of the United Nation
FDLRR	Fully Distributed Linear Reservoir Routing
GCM	Global Circulation Model
GDP	Gross Domestic Product
GIS	Geographical Information System
GLDAS	Global Land Data Assimilation

GPCC	Global Precipitation Climatology Centre
GPCP	Global Precipitation Climatology Project
HBV	Hydrologiska Byråns Vattenbalansavdelning
IFS	Integrated Forecast System
IPSL	Institut Pierre Simon Laplace
IR	Infrared
JRC	Joint Research Centre
LRR	Linear Reservoir Routing
MSWEP	Multi-Source Weighted Ensemble Precipitation
MW	Microwave
NASA	National Aeronautics and Space Administration
NED	National Elevation Dataset
NetCDF	Network Common Data Form
NSCE	Nash-Sutcliffe Coefficient of Efficiency
ORCHIDEE	Organizing Carbon and Hydrology in Dynamic Ecosystems
PERSIANN	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Network
PET	Potential Evapotranspiration
QDLRR	Quasi Distributed Linear Reservoir Routing
QQ	Quantile-Quantile
RCM	Regional Climate Model
RMSQD	Root Mean Square Difference
SCE-UA	Shuffle Complex Evolution University of Arizona
SWAT	Soil and Water Assessment Tool

TMPA	Multi-satellite Precipitation Analysis
TRMM	Tropical Rainfall Measuring Mission
TSC	Temporal Spatial Constant
TSV	Temporal Spatial Variable
UNEP	United Nations Environment Programme
UniK	Universitat Kassel
UniUt	Universiteit Utrecht
USGS	United States Geological Survey
WaterGAP	Water Global Assessment and Prognosis
WCI	Water Cycle Integrator
WCRP	World Climate Research Programme
WMO	World Meteorological Organization
WRR	Water Resources Reanalysis
WWAP	World Water Assessment Program

Chapter One

1. General Introduction and Background

1.1. Background and Motivation

Earth is known as the "Blue Planet" because 71 percent of the Earth's surface is covered with water. Water also exists below land surface and as water vapor in the air. The earth has an abundance of water, but unfortunately, only a small percentage (about 0.3 percent), is even usable by humans and the other 99.7 percent is in the oceans, soils, icecaps, and floating in the atmosphere [*Gleick and White*, 1993]. This limited freshwater resources of the earth are distributed into hydrological storages as glaciers, groundwater, freshwater lakes and wetlands, soil moisture, atmospheric water and river waters [*Shiklomanov and Rodda*, 2003].

Most of the water used by humans comes from rivers (freshwater) and the spatio-temporal distribution of freshwater is very much non-uniform across the globe. The world's rivers carry on average 40% of total land precipitation to oceans or inland sinks [*Gerten et al.*, 2008]. This river discharge exhibits pronounced inter-annual variability imparted by climate variability, and recent climatic changes have significantly altered its amount and seasonality in many regions. In addition, discharge is influenced by human activities such as irrigation, river regulation, dam construction, and land cover changes [*Gerten et al.*, 2008; *Kundzewicz et al.*, 2008; *Vörösmarty and Sahagian*, 2000].

The major share of the total water withdrawals and consumptions pertaining to the agriculture sector (about 70%) followed by industrial and municipal sectors. The world water withdrawals have increased over 7 times during the last century, i.e., from 578 km³yr⁻¹ in the year 1900 to about 3,788 km³yr⁻¹ in the year 1995 [*WWAP*, 2006]. This trend is projected to continue in future,

though with comparatively lower rates. As a consequence, the freshwater resources of the world are under ever-increasing pressure due to escalating demands.

In recent decades, with the explosive growth of world's population and rapid economic development, the shortage of freshwater has become a serious problem worldwide [Vörösmarty *et al.*, 2000]. This indicates that the world is facing a serious freshwater crisis and balancing freshwater uses for humans and nature is examined as the major challenge of this century. All the signs suggest that it is getting worse and will continue to do so, unless corrective action is taken.

This is especially true for developing countries, particularly in Ethiopia, where the need to enhance agricultural productivity faces diminishing freshwater supplies for agriculture owing to rapid industrialization and ever-increasing municipal needs. In Ethiopia, the uneven spatial and temporal occurrence and distribution of the freshwater resources and data scarcity among others are considered as the major factors affecting the development and management of water sector of the country [Allen *et al.*, 1998].

For this reason, under the condition of water scarcity and competing water uses, improved knowledge of basin-wide hydrology and resource (data) availability are pivotal to instruct informed policy formulation and sustainable development of the water sector. Across the globe there is thus an increasing need for water resource availability assessment at basin scale for the water resources planning and management in a country using hydrological models.

Hydrological models provide an important tool for re-constructing and understanding past hydrological variability, and for quantifying future hydrological changes for water resources planning and managements for various applications in a basin scale. Setting up hydrological models need accurate and long records of observed data with different temporal and spatial scales.

Especially, fully distributed models are capable of capturing the spatial distribution of input variables including meteorological conditions (rainfall, potential evapotranspiration etc.) and physical parameters (elevation (DEM) and its derived outputs of drainage network, slope, flow accumulation and flow direction). So, distributed models are data intensive they need quality data and they require greater simulation and calibration time. The model will subdivide a sub-basin area based on a particular grid size to capture spatial and temporal variability and then route flows through the sub-basin from cell to cell. Still, distributed hydrological models have uncertainty problem to simulate the observed flow. Uncertainty in modeling results is investigated with the change in data length and the accuracy of modeling deteriorated with the decrease in data length and that the best-fitting copula model depended on the data length [Tong *et al.*, 2015]. These show that length and quality of hydro-meteorological data sets are vital for the sustainable water resource assessment and applications. However, the availability and quality of climate data varies from region to region, and observations are sparse and irregularly distributed in many areas [El-Sadek *et al.*, 2011; Haberlandt and Kite, 1998; Kanamaru and Kanamitsu, 2007]. Especially in Africa in which the rainfall gauging stations are sparse and irregularly distributed and acquiring accurate and consistent hydro-meteorological observation data set in different temporal and spatial scales is a challenging task.

To overcome the data limitation problem for hydrological modeling and for water resources applications, different researchers use remotely-sensed and long-term reanalysis spatial and temporal scales of global data (precipitation, temperature, solar radiation, humidity and wind speed) for instance [G. Balsamo *et al.*, 2015; Habib *et al.*, 2014a; Sahu *et al.*, 2017]. Alternatively, recently Water Resources Reanalysis (WRR) river runoff products are produced at a global scale using global hydrological model at different temporal and spatial scales. Reanalysis studies, that

merge climate model predictions and observation data, are used to develop long-term historical spatially and temporally distributed climatic variables such as precipitation, temperature, solar radiation, humidity and wind speed [*Bromwich and Fogt, 2004; Overpeck et al., 2011*]. The global products are produced for different spatial and temporal scales for various purposes of water resources applications.

Therefore, high-resolution precipitation and global river runoff (WRR) products have major role on the hydrological analysis and water resources application in data limited regions as Africa. However, high-resolution satellite and reanalysis precipitation products have their own errors and different performance on estimation of the actual measured rainfall on the ground to simulate the streamflow for hydrological analysis, as well as the reanalysis global river runoff (WRR) products have similar setback in estimation of the observed streamflow data.

Numerous hydrological performance researches are conducted on various global satellite precipitation products across the globe for instance [*H Beck et al., 2017; Mei et al., 2014; Seyyedi et al., 2015a*]. Mei et al., 2014 investigated error characteristics of six quasi-global satellite precipitation products and their error propagation in flow simulations for a range of mountainous basin scales (255–6,967 km²) for different periods (May–August and September–November) over the Upper Adige river basin (Eastern Italian Alps). Seyyedi et al., 2015a also assessed the error propagation of two global (or near-global) precipitation datasets in terms of flood modelling for a range of basin scales (300–70,000 km²) focusing on multi-year (2002–2011) simulations over a mid-latitude basin (Susquehanna River Basin) in the North-eastern United States. The datasets are the Tropical Rainfall Measuring Mission (TRMM) post-real-time research version 3B42v7 research product and the Global Land Data Assimilation (GLDAS) reanalysis system precipitation dataset. H Beck et al., 2017 evaluated the performance of Multi-Source Weighted Ensemble

Precipitation (MSWEP) version 1.1 using simple conceptual hydrological model Hydrologiska Byråns Vattenbalansavdelning (HBV) for 9011 catchments across the globe.

Similar researches are undertaken in the Blue Nile basin, Ethiopia. Bitew and Gebremichael, 2011 evaluated the hydrological performance of Climate Prediction Center's morphing technique (CMORPH), satellite-only product TRMM Multi-satellite Precipitation Analysis (TMPA) 3B42RT, Satellite-gauge product TRMM TMPA 3B42 and Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) for small size watersheds of Gilgel Abbay (1,656 km²) and Koga (299 km²) in Abbay basin, Ethiopia using a semi distributed hydrological model Soil and Water Assessment Tool (SWAT). Habib et al., 2014b also showed improvements bias corrected CMORPH satellite data on the hydrological performance of Gilgel Abbay (1,656 km²) watershed compared to uncorrected CMORPH product using semi distributed HBV hydrological model.

Most of these studies show to have limitation of considering the effect of temporal and spatial scales. Furthermore, these are carried out on small watersheds and for short simulation periods. Apart from these, evaluation is conducted by independently calibrating each precipitation products as opposed to calibrating using observed data and validate the products with simulation mode.

Lastly, the global reanalysis river runoff (WRR) products which are available with high resolution of long term historical data (from 1980 to 2013) are not evaluated and validated for various local water resources application in the upper Blue Nile basin or elsewhere.

This dissertation research facilitates a better understanding of global remotely sensed and reanalysis products and provides evaluation these products to local water resources application in the upper Blue Nile basin in Ethiopia. In the first part of this research study, application of global

precipitation products (remotely sensed and reanalysis) for local water resources are evaluated using grid based fully distributed hydrological modeling approach. The locally set grid based fully distributed hydrological model Coupled Routing and Excess Storage (CREST) gives the simulated flow, forcing with various global precipitation products to evaluate the hydrological performance of the global precipitation products with different calibration and simulation techniques to improve the performance to simulate the streamflow at different basin scales. In the second part of the research, various long-term water resources reanalysis (WRR) global river runoff products available at earth2Observe data portal are evaluated and bias corrected or validation of global runoff datasets. Spatio-temporally dynamical bias adjustments are used to improve streamflow predictions and produce a new set of bias corrected gridded runoff products consistent with the available grid resolution (0.25°) and daily time scale for a period of 30 years (1980 to 2010).

1.2. Research Questions

The research attempts to assess and address the following basic questions.

1. How much of the error of satellite and reanalysis precipitation products propagates to the simulated flow? Can we characterize the error and update the simulation?
2. Can we describe the local hydrological condition using fully distributed grid based hydrological model and can global precipitation data sets provide a better performance of runoff estimation for different calibration and simulation modes of the products?
3. Are the global datasets of both precipitation and water resources reanalysis (WRR) river runoff ready enough for application to water resources management and development?
4. Can we update and provide a better long-term reanalysis river runoff dataset at un-gauged and gauged with limited data locations that can be generated using any other available data sets?

1.3. Research Objectives

The general objective of the study is evaluation of global water resources data products for local water management application in the upper Blue Nile basin in Ethiopia using grid based distributed hydrological modeling approach (validation of global data sets). Sub-objectives of the study are:

- To evaluate the precipitation products applying different simulation and calibration modes to improve the hydrological performance for small spatial scale of the basin using CREST hydrological model.
- To evaluate the globally available satellite and reanalysis precipitation products for local water resources application at multi-spatial and temporal scales in the upper Blue Nile basin.
- To evaluate the propagation of errors in global water resources data products into local runoff estimated using fully distributed hydrological model.
- To evaluate the global water resources reanalysis (WRR) river runoff products for local water resources application at multi-spatial and temporal scales in the upper Blue Nile basin.
- To develop long term consistent set of gridded high resolution (0.25°) daily streamflow data for local water resources management in the upper Blue Nile basin.

1.4. Thesis Structure

This thesis is composed of seven chapters.

Chapter 1 is the general introduction and background, including the background, problems, research question and objectives.

Chapter 2 is data and methodology, includes general description on study area, in situ hydro-meteorological data, global precipitation and runoff products and study methodology.

Chapter 3 evaluates satellite and reanalysis precipitation products at small scale: a case study of Gilgel Abbay. The study uses different simulation and calibration conditions (modes) to improve the performance of global precipitation products to simulate the observed flow using fully distributed hydrological model (CREST).

Chapter 4 focuses on multi-scale evaluation of globally available satellite and reanalysis precipitation products for water resources management application: a case of three nested watersheds in the upper Blue Nile basin (Abbay), Ethiopia. The study is conducted in the upper Blue Nile basin over multi-scales (1,656–199,812 km²) focusing on multi-year (2000–2012) daily simulation and analysis using CREST model.

Chapter 5 focus on the performance evaluation of two version earthH2Observe Water Resources Reanalysis (WRR) river runoff products, WRR1 (seven products) used WFDEI forcing data available at 0.5° and WRR2 (four products) used MSWEP forcing data available at 0.25°, in terms of flow variability characterization in the upper Blue Nile basin. These WRR runoff products are compared to in situ observed runoff data and contrasted to runoff simulated using a locally calibrated distributed hydrological model (CREST) forced with in situ local as well as the global precipitation dataset (MSWEP) used in WRR for multi-spatial and temporal scales.

Chapter 6 focuses on improving the quality and availability of local river runoff data applying dynamical bias correction schemes, using globally accessible water resources reanalysis river runoff product (WRR) of Universitat Kassel available at 0.25° grid resolution for long-term daily data from 1980 to 2010 for data scarce region of the upper Blue Nile basin, Ethiopia.

Chapter 7 the main results of this thesis are summarized, and conclusion and recommendations are presented.

Chapter Two

2. Data and Study Methodology

2.1. Description of Study Area

Upper Blue Nile (locally known as Abbay) is one of the most important tributaries of the Nile River as it contributes about 60% of the annual total Nile discharge [Sutcliffe and Parks, 1999]. The source of the Blue Nile is the Little Abbay (locally known as Gilgel Abbay as used in this research) and originates from the Ethiopian highlands upstream of Lake Tana. Gilgel Abbay along other many tributaries leave the Lake as Blue Nile river at Bahir Dar town and flows to the southeast through a series of cataracts crosses the main Addis Ababa to Bahir Dar road at Kessie station located near the new Hidase (Renaissance) Bridge. It continues to flow to the northwest direction and to Sudan (measured at Border or Eldiem station). Along the way, the river is joined by several tributaries draining a large area of highlands from the western Ethiopia (Figure 1). The elevation of the basin varies greatly from over 4176 m in the headwaters of some tributaries to 503 m at the foot of the plateau. The flow of upper Blue Nile and its tributaries are marked by a mono-modal flow pattern with peak flow located in July-August season [Elshamy *et al.*, 2009].

The basin has an average annual discharge of about 49.4 Billion Cubic Meter, measured at Sudan border (BCM). It is the most important basin in the country by most criteria as it contributes about 45% of the countries surface water resources, accommodates 25% of the population, 20% of the landmass, 40% of the nations agricultural product and most of the hydropower and a significant portion of irrigation potential of the country [Erkossa *et al.*, 2009]. The sixteen sub-basins that are primary tributaries of Abbay basin are Rahad, Tana, Besheilo, Dinder, Welaka, Jemma, Muger, Guder, Fincha, Anger, Didessa, Dabus, North Gojam, South Gojam, Wombera and Beles (Figure

1). These sub-basins are characterized by a highly seasonal rainfall pattern with most of the rainfall falling in four months (June to September – JJAS) with a peak in July or August. The mean annual rainfall for the 1961–1990 period amounts to a little over 1200 mm, of which more than 70% fall during those four months. More than 80% of annual flow in the Blue Nile is also contributed from those summer months concentrated between June and September [Elshamy *et al.*, 2009; Setegn *et al.*, 2010].

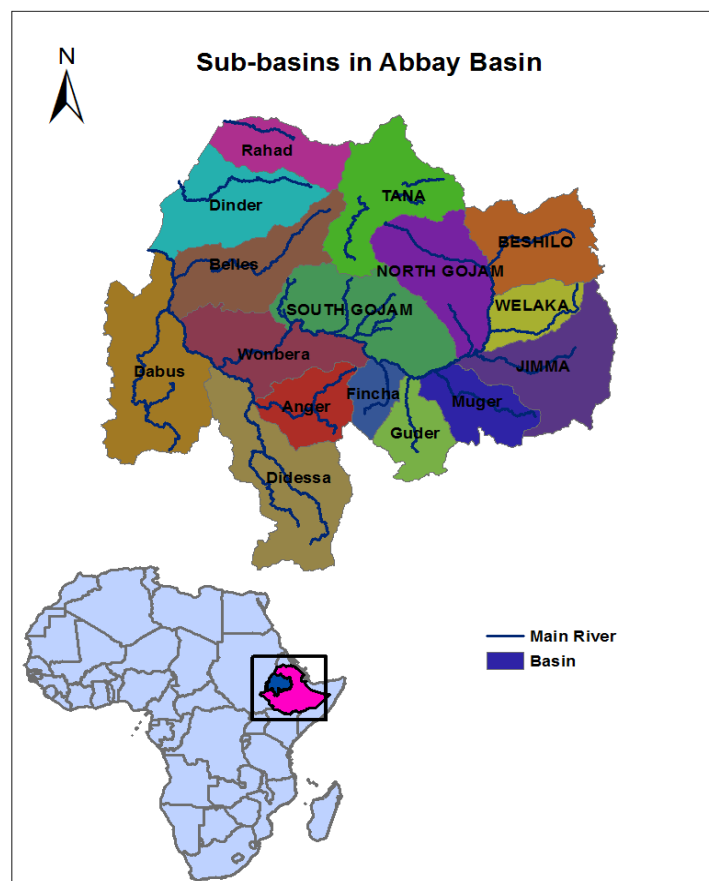


Figure 1. Location of sub-basins in the upper Blue Nile (Abbay) basin.

2.2. In Situ Hydro-Meteorological Data

The daily rainfall historical data are collected from Ethiopian Meteorological Agency. A total of 89 in situ rainfall measuring stations that have good quality and record length (1980 to 2012) are

used for the study. Spatial interpolation of rainfall data is of great importance for hydrological modeling especially for the distributed hydrological models. Geostatistical methods (kriging) are widely applied in spatial interpolation from point measurement to continuous surfaces [Ly *et al.*, 2011]. For this study the available 89 stations with high quality and better length of records are spatially interpolated using Kriging method in Matlab for daily rainfall on 0.25° regular grid resolution in the upper Blue Nile (Abbay) basin area.

Spatial interpolation is generally carried out by estimating a regionalized value at un-sampled points from a weight of observed regionalized values. In this study, points refer to the centers of 25° regular grids in the basin area. The general Kriging formula for spatial interpolation is as follows:

$$Z_g = \sum_{i=1}^{ns} \lambda_i Z_{si} \quad (1)$$

where Z_g is the interpolated value at point g , Z_{si} is the observed points (rain gauges) and $\lambda = \lambda_i$ is the weight contributing to the interpolation.

The station river runoff data set obtained from the Ethiopian Ministry of Water, Irrigation and Electricity used to evaluate the global grid-based reanalysis runoff data and serve as reference to estimate the bias factor of the corresponding pixel of the gridded global data and for the simulation.

To reflect the importance of developing a consistent set of hydrological runoff data in the upper Blue Nile basin, the accuracy and consistency of some selected stations are discussed.

i) Great Anger and little Anger

The two streamflow stations located in the Abbay (upper Blue Nile basin) are measured from two different gauging stations. As the name indicates that great Anger (near to Nekemete, 36°31'E and 9°26'N) is located in the downstream side of the little Anger gauging station (at Gutin, 36°35'E and 9°30'N) including the area of the little Anger watershed. The watershed areas of the two stations of great and little Anger are 4,674 km² and 3,742 km² respectively. The mean daily streamflow of the years from 2000 to 2010 is plotted and shown in Figure 2 to demonstrate the dry season hydrograph of the two stations. From the two stations, the great Anger which is positioned on the downstream side of the little Anger is expected to present a larger magnitude of streamflow in both dry and wet seasons. Nevertheless, the little Anger shows high flow magnitude during the dry season relative to the great Anger flow. Regardless of the sources of the error, these inaccuracies are concerning to planning and design of water resources projects in the area.

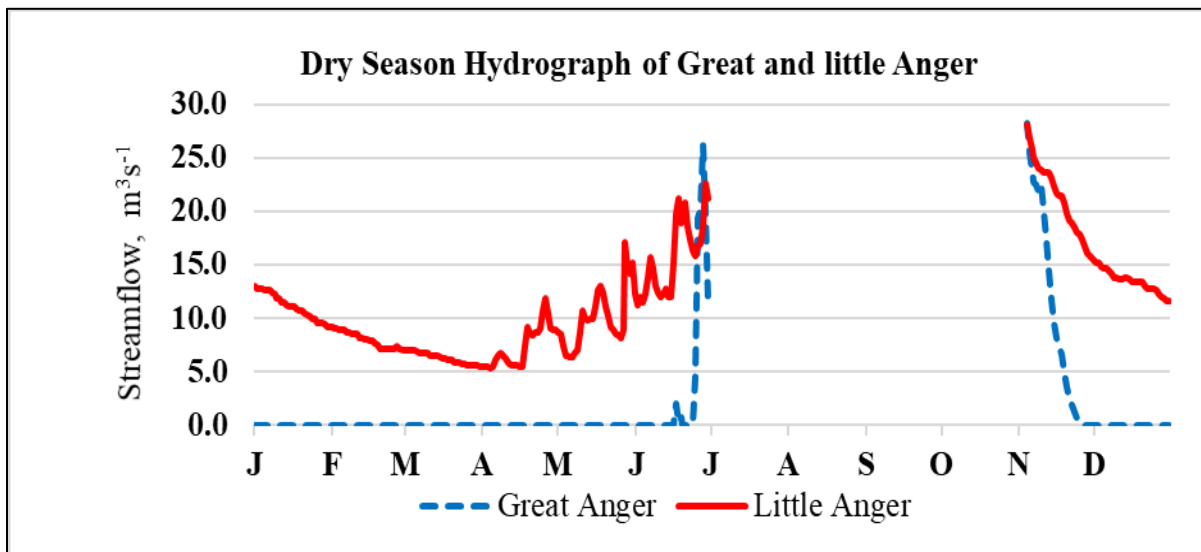


Figure 2. The daily mean Hydrograph of great and little Anger in the upper Blue Nile basin.

ii) Jema

Jema is one of the tributaries of Abbay basin located at the East-Westward of the basin in which the gauging station is located near to Lemi at 38°54'E and 9°55'N with a watershed area of 5,412km². The flow analysis of the river shows that the flow magnitude displays abrupt change after the year of 2002. The river flow analysis is carried out between 1990 to 2010 and the mean daily flow before 2002 (1990 to 2002) and after 2002 (from 2003 to 2010) are plotted as shown in Figure 3. The figure shows that there is shifting between the two mean flows after and before the year of 2002, that might be ascribed to the rating curve deviations to change the measured staff reading in to flow, or the staff might be moved from the original site due to different reasons.

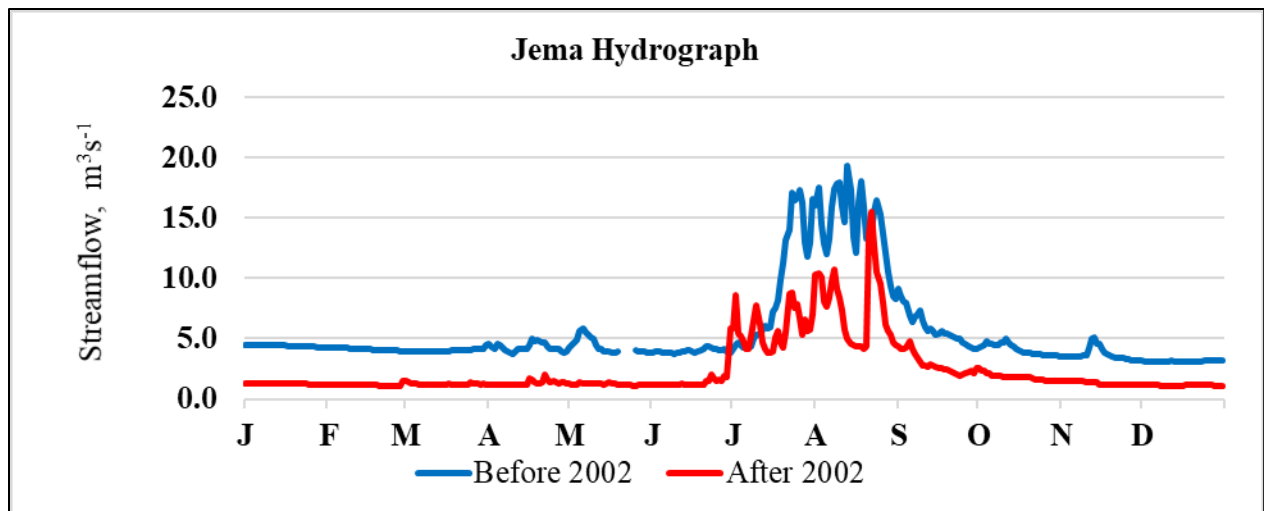


Figure 3. The daily mean hydrograph of Jema watershed in the upper Blue Nile basin.

iii) Gilgel Abbay

Gilgel Abbay watershed, which has a drainage area of 1,656 km², is located in Abbay basin, in the north part which is the major tributary of Lake Tana. The streamflow gauging station is located near to Merawi at 37°02'E and 11°22'N. The monthly mean streamflow from 1980 to 2012 of the dry seasons (January, February, March and April) are plotted. The visual inspection of the

hydrograph shows that the baseflow record of Gilgel Abbay station shows an abrupt increase in late 2005 as shown in Figure 4. This is due to the gauging staff is moved by hundreds of meters downstream of the original site towards the end of 2005 due to road construction. When stations move, rating curves needed to be updated as the velocity and cross section profile changes dramatically. This is no any indication of revising the rating curve.

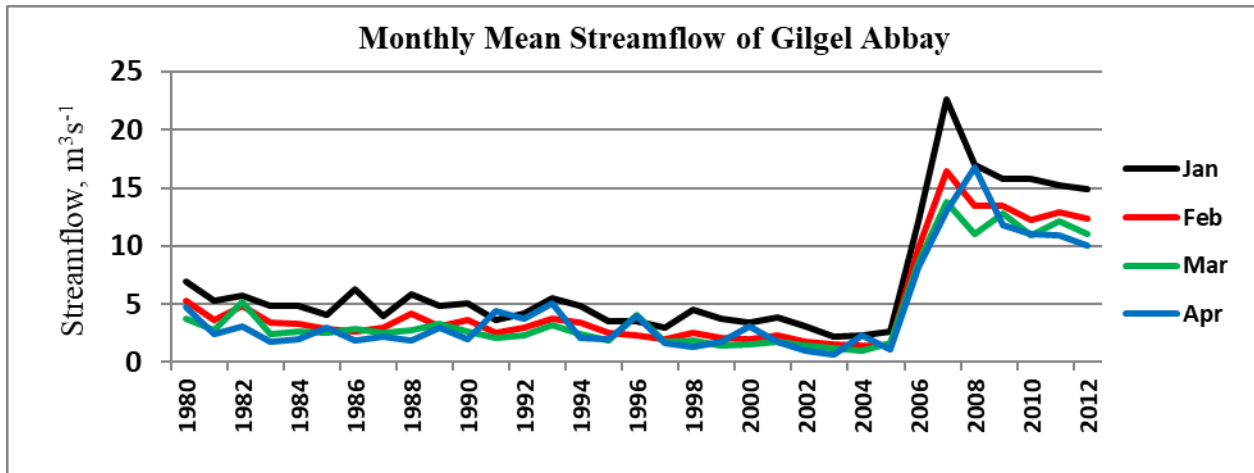


Figure 4. Dry season monthly mean streamflow of Gilgel Abbay watershed in the upper Blue Nile basin.

iv) Kessie Station

Abbay at Kessie gauging station is located at the mid of the Abbay basin, which has a drainage area of 65,784 km², in the upper Blue Nile basin in which the station is located on the main road from Addis Ababa to Bahir Dar. The station flow is computed using two rating curves of the recent one (updated rating curve) and the past (former rating curve). The former rating curve is used to convert the data before 2005 and the recent rating curve is used after the year of 2000. The two rating curves have the common year from 2000 to 2005 that helps to compare the flows that are converted using the recent and former rating curves as shown in Figure 5. The monthly streamflow

hydrograph visual assessment of Figure 5 shows that the river flow converted from the former rating curve shows peak discharge almost twice higher than the flow converted using the updated rating curve. From comparison with downstream stations, it appears the data generated using the updated rating curve is more reliable. For this study the updated rating curve converted data is used for the simulation and comparison.

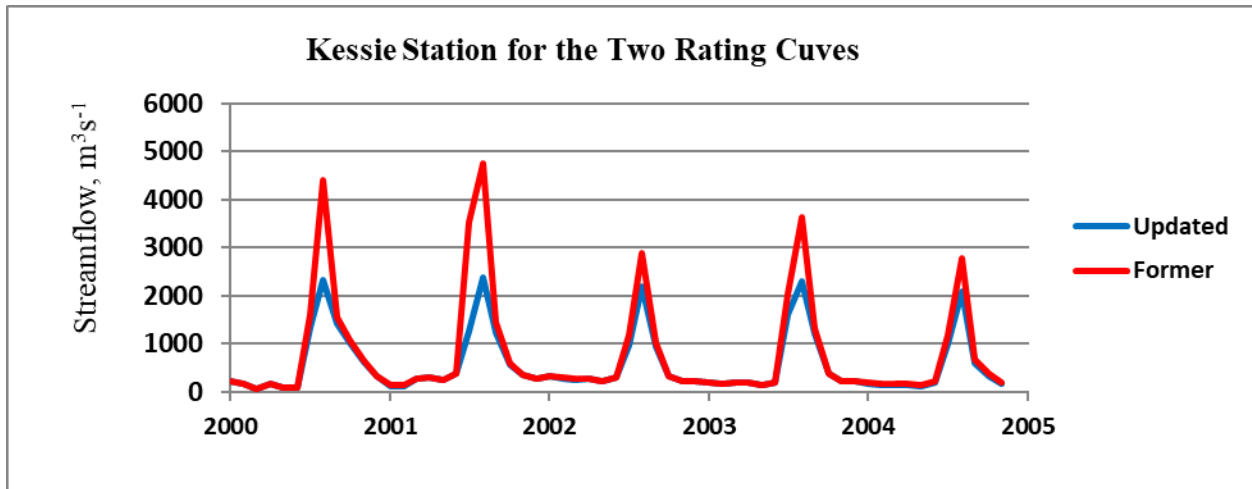


Figure 5. Monthly streamflow of Kessie for the updated and former rating curves.

v) Flow stations at Boarder (Eldiem)

Two stations that are located on the border of Ethiopia and Sudan are operated and managed by Ethiopian and Sudanese governments respectively. The two stations use two different rating curves to convert the staff reading in to streamflow magnitudes. For comparison the two stations monthly flow are plotted and shown in Figure 6. The two stations show different river flow magnitudes for the nearly the same location that should have flows of the nearly the same magnitudes, however, the Ethiopian station flow shows the high peak magnitude relative to the Sudanese station flow. This demonstrates that taking data from different institutes might give a different evaluation performance for specific gauging station of Abbay at Eldiem attributable to that the streamflows

have different recordings. For this study the streamflow data of Eldiem is taken from Sudanese station, since the daily streamflow data is not available from the Ministry of Water, Irrigation and Electricity of Ethiopia.

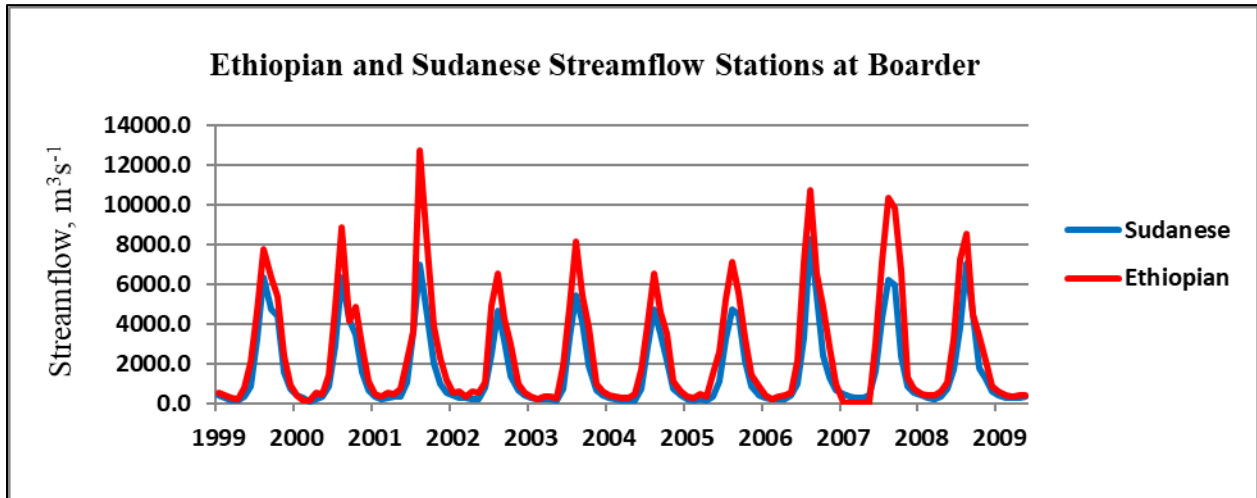


Figure 6. Monthly streamflow hydrograph of Eldiem for Ethiopian and Sudanese stations.

The above observed inaccuracies and inconsistencies indicate the importance of understanding the data before using for water resources application. A slight change in the flow pattern may significantly change the management and design aspects of water resources infrastructure. Furthermore, it undermines our scientific understanding of long term environmental and climate change impacts. The measurement inaccuracies in streamflow is not expected to improve any sooner and seeking alternative sources of data is critical for continued planning, design and management of water resources in Ethiopia. Subsequent sections present the global remotely sensed and reanalysis precipitation and runoff products evaluated as alternative sources of data for local use in the upper Blue Nile Basin watersheds.

2.3. Global Precipitation Products

In this study satellite and reanalysis precipitation products with daily temporal scale are considered for hydrological evaluation of daily simulation flow for different basin scale in the upper Blue Nile basin. Satellite precipitation products of gauge-adjusted CMORPH and TRMM TMPA 3B42v7 with a resolution of (Daily, 0.25°) are considered.

Climate Prediction Centre's morphing technique (CMORPH) products are derived using the morphing technique that uses precipitation estimates derived from passive microwave (MW) observations and propagates these features in space using motion vectors derived from half-hourly geostationary satellite infrared data [Joyce *et al.*, 2004].

Tropical Rainfall Measuring Mission (TRMM) Multi Satellite Precipitation Analysis (TMPA) post-real time research version product 3B42v7 uses MW data to calibrate the infrared(IR)-derived estimates and creates estimates that contain MW-derived rainfall estimates when and where MW data are available and the calibrated IR estimates where MW data are not available [Huffman *et al.*, 2007].

Like satellite precipitation products, reanalysis precipitation products are considered in the hydrological evaluation in this study. Reanalysis data provide a reconstruction of the historical global atmospheric circulation patterns through assimilation of available observation data into Global Climate Model (GCM) predictions. Thus, reanalysis data are widely considered to be a robust proxy of historic atmospheric observations that reproduce observed climate dynamics and are therefore useful for climate application studies where there are inadequate meteorological input data [Hwang *et al.*, 2013]. Reanalysis studies, that merge climate model predictions and observation data, are undertaken to develop long-term historical spatially and temporally distributed climatic variables such as precipitation, temperature, solar radiation, humidity and wind

speed. The reanalysis precipitation products that are considered in this study are briefly described below.

ECMWF (European Centre for Medium-Range Weather Forecasts) is one of the outputs of forcing reanalysis version-1 global precipitation data with coarser resolution of 0.5° spatial and with daily temporal resolutions considered for the simulation of streamflow [Andersson, 2015](<http://www.ecmwf.int/en/research/climate-reanalysis/era-20c>).

ERA-Interim (ERA-I) is reanalysis precipitation product with 0.25° spatial scales. ERA-Interim/Land is a global land surface reanalysis dataset covering the period 1979–2010. It describes the evolution of soil moisture, soil temperature and snowpack. ERA-Interim/Land is the result of a single 32-year simulation with the latest ECMWF (European Centre for Medium-Range Weather Forecasts) land surface model driven by meteorological forcing from the ERA-Interim atmospheric reanalysis and precipitation adjustments based on monthly GPCP v2.1 (Global Precipitation Climatology Project). The horizontal resolution is about 80 km and the time frequency is 3-hourly [G. Balsamo *et al.*, 2015; Dee *et al.*, 2011]. For this study the 3-hourly precipitation data are aggregated into daily for the daily flow computation.

Global Precipitation Climatology Centre (GPCC) is established in 1989 on request of the World Meteorological Organization (WMO). It is operated by Deutscher Wetter Dienst (DWD, National Meteorological Service of Germany) as a German contribution to the World Climate Research Programme (WCRP). GPCC's new global precipitation climatology V.2015 (available in $2.5^\circ \times 2.5^\circ$, $1.0^\circ \times 1.0^\circ$, $0.5^\circ \times 0.5^\circ$, and $0.25^\circ \times 0.25^\circ$ resolution) [Schneider *U. et al.*, 2015]. For the analysis of this study the finer with 0.25° spatial and daily temporal scale is considered.

Multi Source Weighted Ensemble Precipitation (MSWEP) retrospective is a new fully global precipitation dataset (1979–2015) with a high 3-hourly temporal and 0.25° spatial resolutions [H Beck *et al.*, 2017]. The dataset is unique in that it takes advantage of a wide range of data sources, including gauges, satellites, and atmospheric reanalysis models, to obtain the best possible precipitation estimates at global scale. For this study the 3-hourly precipitation data are aggregated into a daily temporal scale for the simulation of daily flow. The spatial distribution of the global precipitation products over the upper Blue Nile is shown in Appendix A.

2.4. Global River Runoff Products

A consistent global reanalysis dataset of water resources, robust and long enough to capture current climate variability and provide better insight into the aforementioned pressures, is currently lacking. At the same time, many tools are available, as well as a multiple of Earth Observation (EO) at a global scale, suitable for supporting water resources quantification if historically coupled and integrated. The accurate mapping and estimation of global water resources remains a major challenge. It requires the use of many sources of earth observations (such as satellite and ground based remote sensing in situ measurements, vertical profiles, etc.), combining with state-of-art earth system modelling components that are developed for hydrometeorological and environmental applications. Together these enables the provisions of continuous maps of present and past conditions. Characterizing not only the global water resources status through all its compartments (soil, snow, inland water bodies, rivers), but also tends over time and the severity of water-cycle extremes.

earth2Observe integrates available earth observations from different satellite missions, in situ datasets from various sources, and state-of-art models to construct a consistent and robust global Water Resources Reanalysis (WRR) dataset of sufficient length (at least 30 years).

Water resources reanalysis river runoff (WRR1, and WRR2) products with daily and monthly temporal resolutions are taken from the earth2Observe data portal (<https://wci.earth2observe.eu/portal/>). The reanalysis products that are considered in this study and the global multi models that generates the products are described below.

The ORCHIDEE (Organizing Carbon and Hydrology in Dynamic Ecosystems) land surface model developed by **CNRS** (Centre National de la Recherche Scientifique) is a mechanistic dynamic global vegetation model [*Krinner et al.*, 2005] that is the part of IPSL (Institut Pierre Simon Laplace) Earth System Model. It is composed of three main modules [*Ducoudré et al.*, 1993] that calculate energy, momentum and hydrological exchanges between soil, vegetation and atmosphere as well as photosynthesis, at a half-hourly time step. The ORCHIDEE land surface model is recently updated to improve the representation of high-latitude environments. The model now includes improved soil thermodynamics and the representation of permafrost physical processes (soil thawing and freezing), as well as a new snow model to improve the representation of the seasonal evolution of the snow pack and the resulting insulation effects [*Dantec-Nédélec et al.*, 2017]. Available only at WRR1 (0.5°) resolution with daily and monthly temporal scales in the data portal of earth2Observe.

The Australian Water Resources Assessment System Landscape model (AWRA-L) developed by **CSIRO** (Commonwealth Scientific and Industrial Research Organization). CSIRO is a grid based distributed biophysical model that simulates water stores and flows in the vegetation, soil and local catchment groundwater systems [*van Dijk and Renzullo*, 2011]. AWRA-L aims to produce interpretable water balance component estimates covering all of Australia, and as much as possible agree with water balance observations, including point gauging data and satellite observations

[Peña-Arancibia *et al.*, 2011]. Available only at WRR1 (0.5°) resolution with daily and monthly temporal scales in the data portal of earth2Observe.

The Tiled **ECMWF** (European Centre for Medium-Range Weather Forecasts) Scheme for Surface Exchanges over Land (TESSEL) is used operationally in the Integrated Forecast System (IFS) for describing the evolution of soil, vegetation, and snow over the continents at diverse spatial resolutions [van den Hurk and Viterbo, 2003]. A revised land surface hydrology (HTESSEL) is introduced in the ECMWF operational model to address shortcomings of the land surface scheme, specifically the lack of surface runoff and the choice of a global uniform soil texture. New infiltration and runoff schemes are introduced with a dependency on the soil texture and standard deviation of orography [Gianpaolo Balsamo *et al.*, 2009]. Available at WRR1 (0.5°) with daily and monthly temporal resolutions, and WRR2 (0.25°) with only daily temporal resolution.

The LISFLOOD model is a hydrological rainfall-runoff model that simulates the hydrological processes that occur in a catchment. LISFLOOD is developed by the floods group of the Natural Hazards Project of the Joint Research Centre (**JRC**) of the European Commission. The specific development objective is to produce a spatially distributed and physically based model that can be used in large and trans-national catchments for a variety of applications, including: flood forecasting, assessing the effects of river regulation measures, the effects of land-use change, and the effects of climate change [Burek *et al.*, 2013]. Available at WRR1 (0.5°) with daily and monthly temporal resolutions, and WRR2 (0.25°) with only daily temporal resolution.

The SURFEX model developed by Meteo France (**MeteoFr**) is composed of various physical models for natural land surface, urbanized areas, lakes and oceans [Le Moigne, 2009]. It also simulates chemistry and aerosols surface processes and can be used for assimilation of surface and near surface variables. SURFEX has its own initialization procedures and can be used in

standalone mode and coupled to an atmospheric model. Available at WRR1 (0.5°) with daily and monthly temporal resolutions, and WRR2 (0.25°) with only daily temporal resolution.

The University of Kassel (Universitat Kassel, **UniK**) WaterGAP3 (Water Global Assessment and Prognosis) is a global model of water availability and water use, is developed to assess the current water resources situation and to estimate the impact of global change on the problem of water scarcity [Alcamo J, 2000; Doell et al., 2001]. WaterGAP comprises of a hydrological module, used to simulate water storages and flows, and a water-use module, which looks at consumption from different economic sectors, including a sub-model for an assessment of global irrigation requirements. The model is used in assessing the impact of global developments (e.g. climate change on water availability and water demand; determining different water-stress conditions of different regions) (<http://www.liaise-kit.eu/model/water-global-assessment-and-prognosis>).

Available at WRR1 (0.5°) with daily and monthly temporal resolutions, and WRR2 (0.25°) with only daily temporal resolution. The model computes time-series of fast-surface and subsurface runoff, groundwater recharge and river discharge as well as storage variations of water in canopy, snow, soil, groundwater, lakes, wetlands and rivers.

The PCR-GLOBWB is a large-scale hydrological model intended for global to regional studies and developed at the Department of Physical Geography of Utrecht University (**UniUt**, Netherlands). PCR-GLOBWB uses process-based equations to compute moisture storage in two vertically stacked soil layers as well as the water exchange between the soil and the atmosphere and the underlying groundwater reservoir [Petrescu et al., 2010]. Exchange to the atmosphere comprises precipitation, evapotranspiration and snow accumulation and melt, which are all modified by the presence of the canopy and snow cover. The exchange with the underlying groundwater reservoir comprises deep percolation and capillary rise and vertical fluxes

<http://pcraster.geo.uu.nl/projects/applications/pcrglobwb/>). Available only at WRR1 (0.5°) resolution with daily and monthly temporal scales in the data portal of earth2Observe.

These WRR runoff products are not validated for local basin study and not checked the spatial and temporal issues of the reanalysis grid-based product with the original ground based geo-spatial data and its derived products (drainage network, flow accumulation and flow direction) of the local basin. This work points to determine the spatial and temporal issues of different reanalysis grid-based products (river runoff) and to have some notes on the spatial and temporal organization of earth2Observe Water Cycle Integrator (WCI) data portal products for the local basin study of the upper Blue Nile (Abbay) basin, Ethiopia. Three gauging station are considered for the analysis of the spatial and temporal organization of earth2Observe WCI data portal, Gilgel Abbay (at 37°02'E and 11°22'N), Kessie (at 38°07'E and 10°05'N) and Eldiem at boarder (at 34°55'E and 11°14'N).

For the temporal and spatial analysis, the global river runoff products at daily and monthly temporal scales from 1980 to 2013 that are considered for spatial and temporal issues analysis are CSIRO (Commonwealth Scientific and Industrial Research Organization), ECMWF (European Central for Medium-Range Weather Forecasting), Meteo France (MeteoFr), Universitat Kassel (UniK) and Universiteit Utrecht (UniUt). These data are available in the earth2Observe data portal in both WCI and catalog formats. The WCI data portal has access to clip the specific location (Spatial issue) of the local basin study from the global spatial data, as well to clip a range of period (Temporal issue) from the total available period range of 1980 to 2013, which helps to download and handle the reanalysis data considerably. The analysis only focuses on the data that are processed from the WCI data portal, which has the only option to clip the data and to have some notes on the spatial and temporal organization of the WCI data portal products. The analysis of spatial and temporal subjects of different reanalysis products are conducted using clipped data of

upper Blue Nile (Abbay) basin. From the whole available date from 1983 to 2013 of the product, the temporal range from 2000 to 2011, and from the whole global data, the spatial range of 34° left, 40° right, 7.5° bottom, 13° top of selected local basin study is clipped. The observed geo-spatial data of Digital Elevation Model (DEM) and its derived products (flow accumulation, flow direction and drainage network) of the local basin are used for spatial evaluation of the connectivity of series grid flow and flow accumulation of generated grid based global reanalysis river runoff data of the WCI data portal.

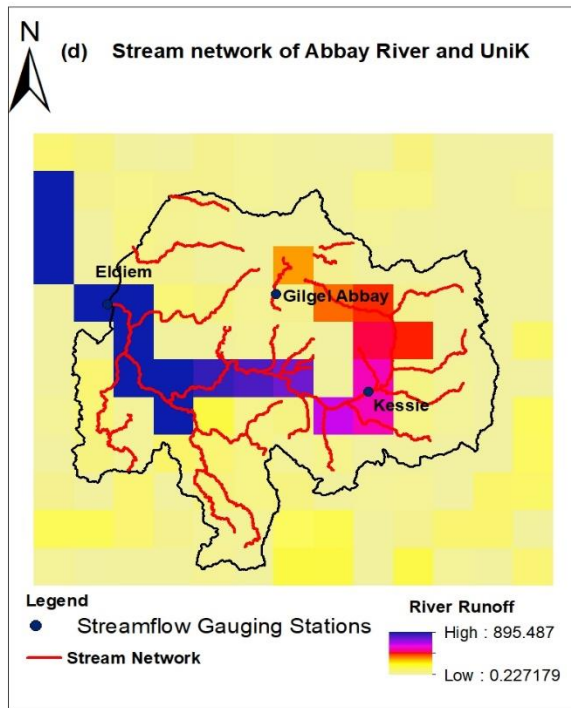
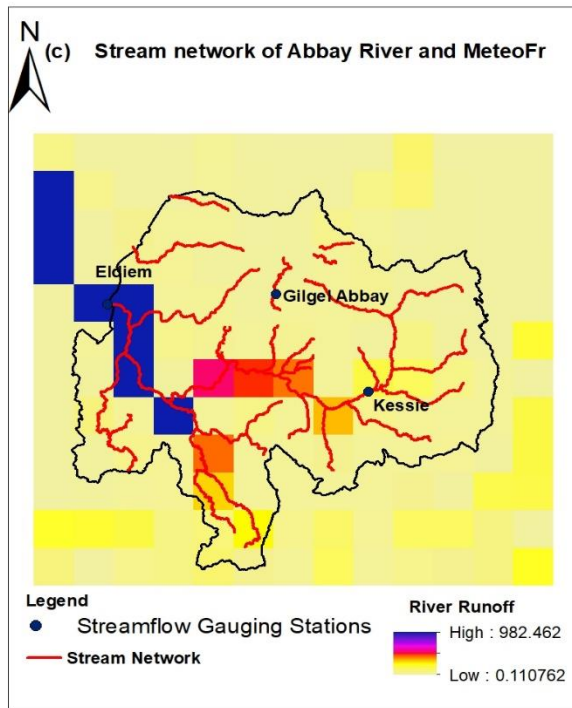
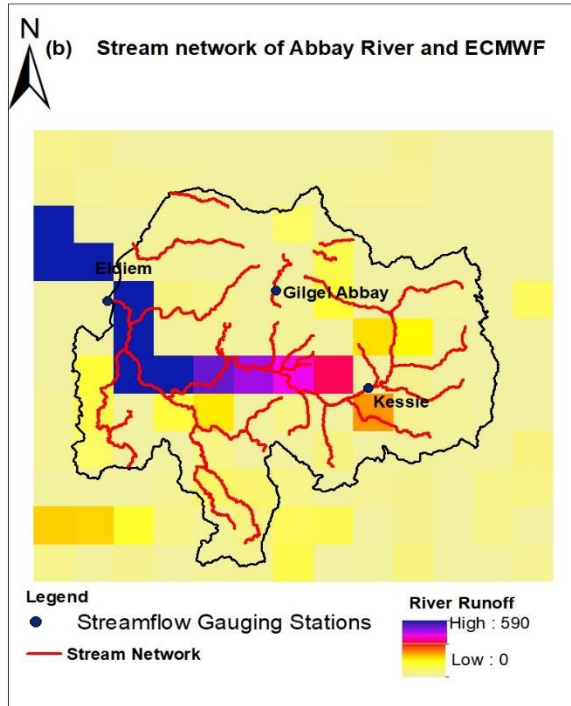
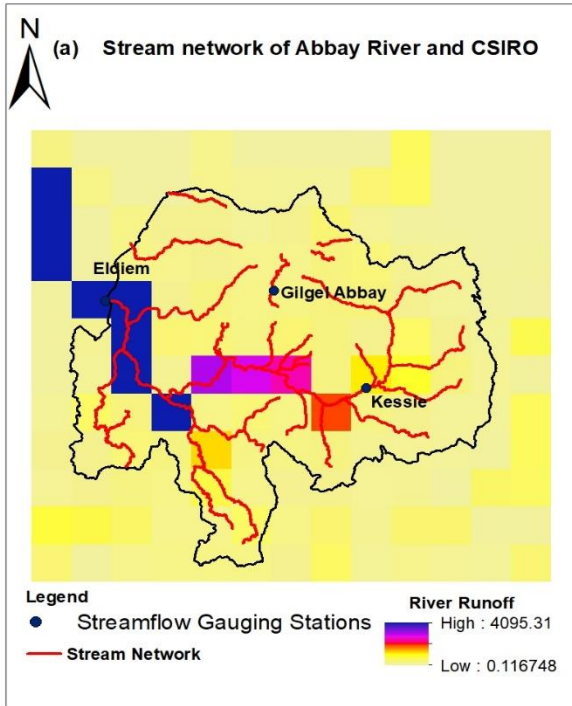
2.4.1. Spatial Issue of Global Runoff Products

The clipped data from WCI data portal of the first round WRR grid based 0.5° resolution of reanalysis river runoff product is spatially located and clipped from 34° left, 7.5° bottom, 40° right, 13° top. Monthly and daily temporal resolutions of available five different products of reanalysis river runoff are considered for the analysis. The spatial maps of the five different WRR1 river runoff clipped products show that each product has different flow accumulation, flow direction and flow connectivity from grid to grid of the local basin study of upper Blue Nile as shown in Figure 7. The flow accumulation of each reanalysis product is overlaid and compared with the streamflow network generated using the existing DEM derived product or the reference stream network of the upper Blue Nile. The overall result of WRR1 products shows that UniK flow accumulation capture very well the reference or the existing river network derived from DEM of the local study of the upper Blue Nile basin. A more colored grid of reanalysis cell indicates that the flow direction is toward the colored cell and shows more flow accumulation. These more colored grids be supposed to trace the existing stream network of the study basin. The result from CSIRO indicates that the flow accumulation tries to get the natural river network very well, particularly at the outlet (Eldiem) and at different tributaries of the river basin. However, it misses

some part of the joining point of the large tributaries of the basin with the mainstream of upper Blue Nile as shown in Figure 7a. The second reanalysis river runoff product is ECMWF, the spatial distribution of flow accumulation of this product has a high mismatch with the reference stream network of the basin (at Eldiem and Kessie), and it has also a grid flow connectivity problem at the upstream of the Kessie gauging station as shown in the Figure 7b. The joining points of large tributaries with the mainstream of the basin are not captured in this reanalysis (ECMWF) river runoff product. The MeteoFr reanalysis product shows that the same spatial distribution of flow accumulation with the CSIRO product. The flow accumulation that is generated using MeteoFr data tries to capture very well the natural stream network of the study basin. However, the same with the CSIRO product there is mismatch at some places of the joining point of the large tributaries with the mainstream of the basin as shown in Figure 7c. The fourth type of the reanalysis river runoff product is UniK, this product capture very well in all aspects of the comparison that have been made in this study. It captures very well each joining point of the tributaries with the mainstream, it has very good grid flow connectivity of each grid flow accumulation and it takes the right streamflow network of the existing streamflow of the local basin study of the upper Blue Nile basin. This product shows better results from the rest reanalysis of monthly and daily river runoff products in terms of capturing the reference stream network as shown in Figure 7d. The last reanalysis product that is considered in this study is the UniUt product. The result from this product shows that, the generated flow accumulation from UniUt tries capture very well the spatial dispersion of the existing river network the same with UniK, but, it has some mismatches at the topographic place of joining points of large tributaries with the mainstream (near to Kessie) as represented in Figure 7e. From the entire reanalysis river runoff products, the UniK product shows better and reliable spatial flow accumulation network, which

matches with the existing river network of the local basin study. On the basis of this manual inspection, the flow accumulation is locally revised as part of this study and corrected.

In general, global runoff products typically do not account the effect of routing flow for reservoirs. The global model river runoff products are not the best applied (used) at the downstream of any reservoir or large-scale diversion for the water resources management application.



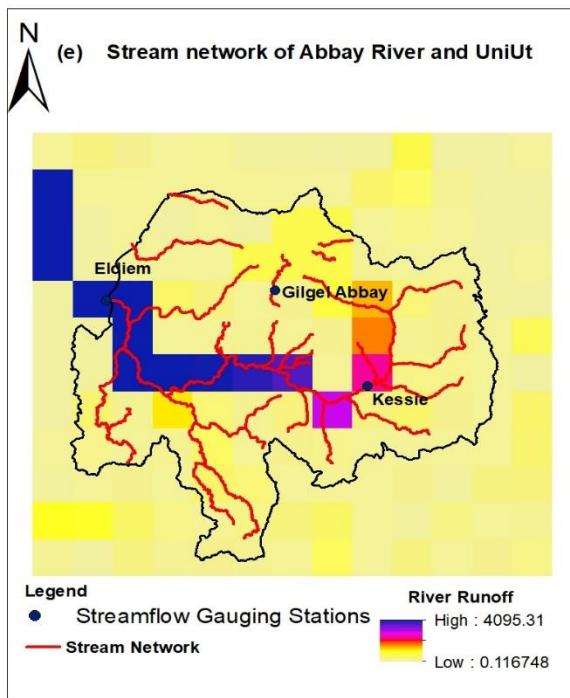


Figure 7. Stream network of the upper Blue Nile (Abbay) basin with different reanalysis river runoff products (WRR1, 0.5°).

2.4.2. Temporal Issue of Global Runoff Products

The WCI data portal has access to clip the specific study time range from the total available date from 1980-01-01 to 2013-12-31. For the temporal issue analysis, we take the monthly and daily river runoff of reanalysis products from 2000-01-01 to 2012-12-31. GIS (Geographical Information System) and Matlab are applied to extract each grid data for the analysis of the temporal issue. Handling and reading the NetCDF (Network Common Data Form) file that is clipped and downloaded from the WCI data portal of the reanalysis products using GIS is very easy to observe the temporal issue for the organization of the data. Since, GIS can directly read the NetCDF file and provides the magnitude of the river runoff and the corresponding date. From this point, it is possible to observe that the initial date that is given to clip the data from the WCI

data portal and the date that is displayed using GIS has one step period difference. To clip the exact initial date of the products in WCI portal, it is quite important to understand that, the date must be lag in one step period (+1). Unless the given date to clip the product and the clipped product initial date shall hold one step period difference, which makes the clipped products will cause one step time difference with the observed data.

Handling the clipped product using Matlab cannot be helpful to identify the initial dates of both clipped product and the date that is given to clip in WCI data portal. Since, Matlab doesn't give and display the corresponding clipped NetCDF product date. Yet it is possible to identify the number of rows, columns and number of periods that are available on the clipped product, which doesn't help and give the exact initial and end date of the clipped products. However, extracting data of each grid cell river discharge data using the number of rows and columns using Matlab can squeeze and give the exact grid values in which the data need to be extracted. To extract the data of each grid cell values using latitude and longitude coordinates applying GIS might give shifted grid or neighbor grid values. Even though, the exact location of the latitude and longitude coordinates are given for GIS to extract the corresponding grid values of specific station location, sometimes it has the possibility to extract and give the neighbor grid values. In general GIS is very helpful to identify the initial and end dates of the clipped data, but not precise to extract each grid values using latitude and longitude coordinates. Matlab has shown better performance to extract the exact grid values using the number of rows and columns of the location of the station, but not to identify the initial and end dates of the clipped reanalysis WCI data portal products.

2.5. Subject for Revision

Issue on Spatial Global Runoff Products

The WRR1 grid based 0.5° spatial resolution of different reanalysis products, clipped from the WCI data portal have their own different spatial distribution of flow accumulation. These reanalysis products are overlaid and compared with the existing streamflow network of the upper Blue Nile basin. The result shows that UniK product of reanalysis river runoff product captures very well the reference streamflow network of the basin in all aspects of comparison, it has good series grid flow connectivity of each grid flow accumulation, it captures very well each joining points of the tributaries with the main stream network and it follows the right streamflow network of the existing flow direction of the upper Blue Nile basin. The products from CSIRO, MeteoFr and UniUt need to have revision on the spatial mismatch at some places of the joining point of the large tributaries and mainstream with the existing river network, but, they have good connection series of accumulated flow grids. ECMWF product needs more revision on the aspects of connectivity of the accumulated flow grid series, missing the right network of the mainstream and joining points of large tributaries from the mainstream of the existing river network. Some grids in the ECMWF reanalysis data shift from the natural network of the mainstream of the local basin, and need shifting the grid value to match with the exact spatial location of the existing mainstream network.

Issue on Temporal Global Runoff Products

The reanalysis clipped product of the WCI data portal regarding on the temporal issue shows that, to clip the exact initial and end dates on the WCI portal, both initial and end dates should be given to the WCI portal one step lag period (+1) of the time period need to be clipped. This is handled easily using GIS, since GIS can read the clipped and downloaded reanalysis river runoff NetCDF and display the magnitude and the corresponding date of the clipped reanalysis product. This temporal issue needs revision on the clipping of the initial and end dates of the products from the

total available time range from 1980-01-01 to 2013-12-31, which needs one step lag period from the initial and end dates of the WRR1 river runoff products.

High resolution WRR runoff products

The second round of the project releases WRR2 reanalysis products, that have spatial resolution of 0.25° with only daily temporal resolution. From the entire WRR1 products, UniK shows the best fit with the existing network of the mainstream of the upper Blue Nile basin. The same with the coarser product of UniK, the high resolution of WRR2 products of UniK product of reanalysis river runoff product also captures very well the reference streamflow network of the basin in all aspects of comparison. It has good series grid flow connectivity of each grid flow accumulation, it captures very well each joining points of the tributaries with the main stream network and it follows the right streamflow network of the existing flow direction of the upper Blue Nile and shown in Figure 8.

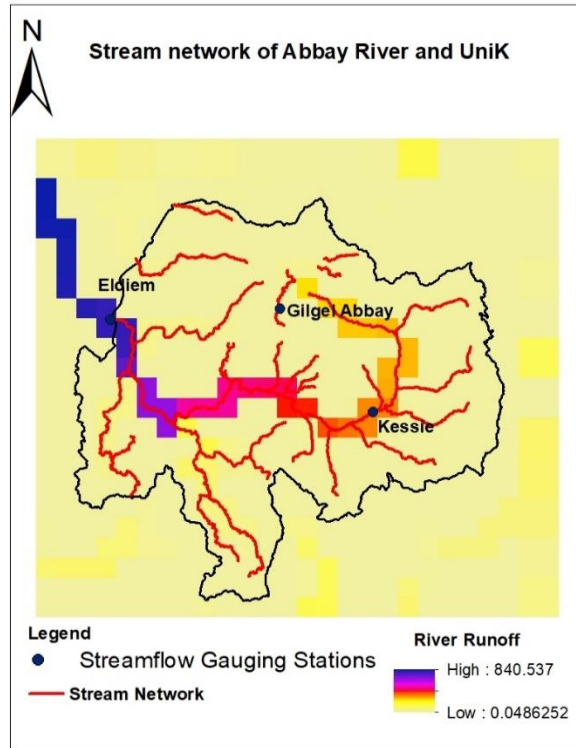


Figure 8. High resolution (0.25°) reanalysis river runoff product of UniK with the existing network of the upper Blue Nile basin.

2.6. Global Potential Evapotranspiration

The grid based (0.5°) with daily temporal scale of reference potential Evapotranspiration (PET, mm/day) that is derived from Penman-Monteith method [Richard G. Allen *et al.*, 1998] is used as an input to setup hydrological model for the hydrologic simulations. The grid PET data is accessed from the earth2Observe data portal (<https://wci.earth2observe.eu/portal/>). The georeferenced PET of the Abbay basin is shown in Figure 9.

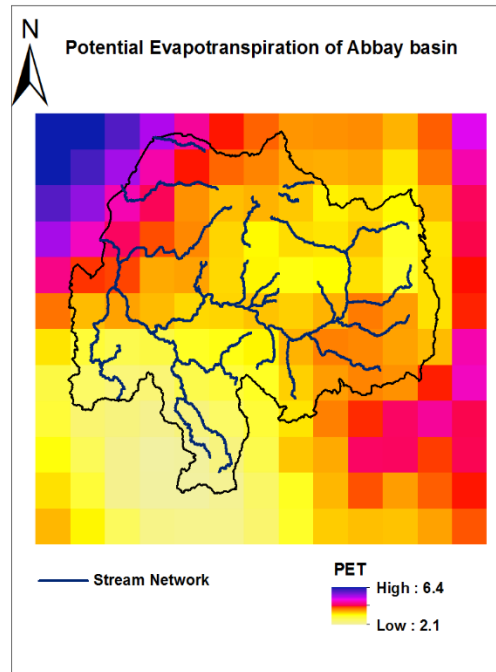


Figure 9. Potential Evapotranspiration derived from Penman-Monteith of the Upper Blue Nile (Abbay) basin.

2.7. CREST Model

2.7.1. Overview of CREST

The Coupled Routing and Excess Storage (CREST) distributed hydrological model is a hybrid modeling strategy developed by the University of Oklahoma (<http://hydro.ou.edu>) and NASA SERVIR Project Team (www.servir.net). The CREST model is initially developed to provide real-time regional and global hydrological prediction by running at fine spatiotemporal resolution with maintaining economical computational cost (<http://eos.ou.edu>). CREST simulates the spatiotemporal variation of water and energy fluxes and storages on distributed grid cells of arbitrary user-defined resolution, which enables multi-scale applications (Figure 10). The scalability of CREST simulations is accomplished through the sub-grid scale representation of soil moisture storage capacity (using a variable infiltration curve), multi-scale runoff generation

processes (using multi-linear reservoirs) and a fully distributed routing scheme (using the fully distributed linear reservoir routing (FDLRR)). The primary water fluxes such as infiltration and routing are physically affected by the geographic variables land surface characteristics (i.e., vegetation, soil type, and etc.). The runoff generation and routing components are coupled; therefore, CREST includes more realistic interactions between lower atmospheric boundary layers, terrestrial surface, and subsurface water than other distributed hydrological models. The above features make CREST applicable at global, regional, and catchment scales [Xinyi Shen and Hong, 2014].

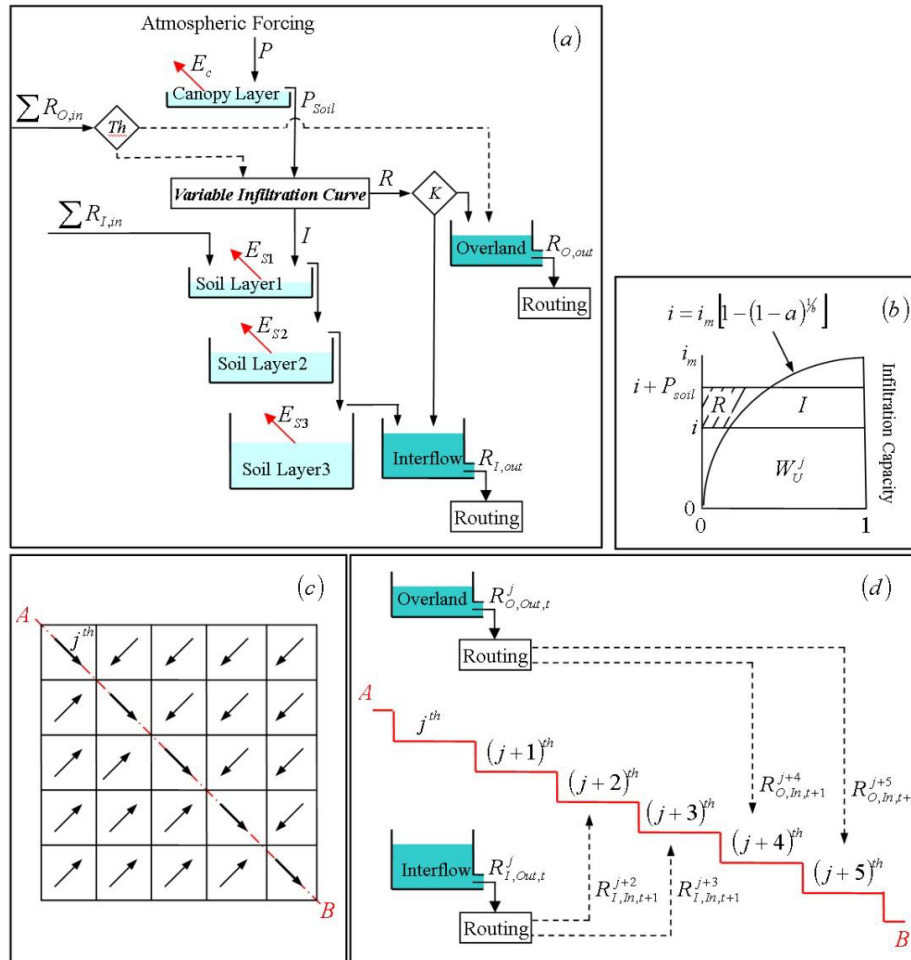


Figure 10. Core components of the CREST model

(a) Vertical profile of a cell including rainfall-runoff generation, evapotranspiration, sub-grid cell routing and feedbacks from routing; (b) Variable infiltration curve of a cell; (c) Plane view of cells and flow directions; and (d) Vertical profile along several cells including sub-grid cell routing, downstream routing, and subsurface runoff redistribution from a cell to its downstream cells.

2.7.2. New in CREST V2.1

The major upgrade is on the routing scheme. The cell-to-cell routing scheme used in previous versions of CREST is a quasi-distributed linear reservoir routing (QDLRR) method which we found problematic to apply to a distributed hydrological model in practice. In CREST v2.1, a fully distributed LRR method (FDLRR) is proposed and used to replace the QDLRR module of CREST. The concept of the QDLRR and the FDLRR is shown in Figure 11. Suppose that in one-time step, water moves from A to D, B to E and C to F and take C as the observation cell. In previous version as shown in Figure 11a, only the water from C to F contributes to the final runoff (discharge) of C while water from A to D and B to E is denied of contribution as if the water jumps over cell C. On the contrary, in CREST v2.1, all these three terms contribute to the runoff of cell C because they either set off from or passed via cell C as shown in Figure 11b [Xinyi Shen and Hong, 2014].

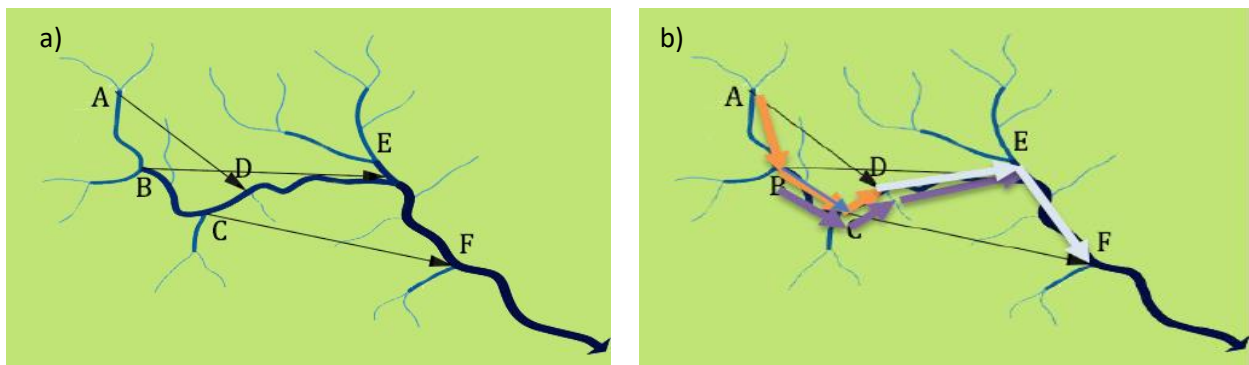


Figure 11. Routing concept (a) Linear reservoir routing (LRR) method and (b) Fully distributed reservoir (DLRR) method

Minor upgrades include: 1) The full vectorization of the computation which boots the efficiency by nearly one order (no routines loop cell in the new version); 2) acceptance of more advanced geographic data formats; 3) Automatic decompression, re-projection, resampling and clipping of the forcing data to accommodate data in different formats, coordinate system and resolution; 4) Adding a flood event mode and 5) Switching on and off a) The feedback mechanism, b) The existence of interflow in channels.

CREST's distinguishing characteristics include:

- Distributed rainfall–runoff generation and cell-to-cell routing;
- Coupled runoff generation and routing via three feedback mechanisms (Sub-grid-scale routing, Downstream routing and Coupling rainfall-runoff generation and routing); and
- Representation of sub-grid cell variability of soil moisture storage capacity and sub-grid cell routing (via linear reservoirs). The coupling between the runoff generation and routing mechanisms allows detailed and realistic treatment of hydrological variables such as soil moisture. Furthermore, the representation of soil moisture variability and routing processes at the sub-grid scale enables the CREST model to be readily scalable to multi-scale modeling research [*Xinyi Shen and Hong, 2014*]. In this study, CREST V.2.1, which involves an updated routing scheme based on a fully distributed linear reservoir (FDLRR) scheme [*Xinyi Shen and Hong, 2014*], has been implemented.

Model Setup

The requirements of data to setup the CREST hydrological model is discussed. Two types of data are required i) the geospatial input data which are static in nature and ii) the hydrometeorological input data which are dynamic time series in nature. The geospatial data include the 1 km by 1km resolution Digital Elevation Model (DEM) extracted from the United States Geological Survey

(USGS) National Elevation Dataset (NED). The DEM is used to derive the geospatial data for CREST model such as slope, flow direction and flow accumulation and drainage area of the case studies. The second input consists of the dynamic hydrometeorological datasets. The precipitation dataset is gridded and geo-referenced to the model grid resolution which is consistent with the satellite and reanalysis resolution of 0.25° . The gridded daily PET data is obtained from earth2Observe data portal and the observed rainfall is gridded and geo-referenced into spatio-temporal daily data using Matlab with 0.25° resolution. Daily time series observed streamflow datasets are used at the outlet of the watersheds to calibrate model parameters.

2.8. Parameters

Many of the parameters in the CREST model can be estimated based on the availability of field survey data, such as soil surveys, land cover maps and vegetation coverage. There are approximately twelve parameters (Table 1) that are much more difficult to estimate from ancillary data and need to be calibrated either manually, automatically, or using combined approaches given observations of rainfall and streamflow [Wang *et al.*, 2011]. As will be elaborated in subsequent chapters, manual calibration of all the CREST model parameters could be time consuming and less practical; therefore, automatic calibration is implemented to estimate the parameters. For automatic calibration, CREST v.2.1, the Shuffle Complex Evolution University of Arizona (SCE-UA) [Duan *et al.*, 1992] is selected as the kernel algorithm in the calibration process. Typically, the procedure involves selection of particles (samples) in the parameter space through the use of competitive evolution schemes, such as the simplex scheme, to reproduce better offspring particles. After generations of evolutions, the population attempts to converge to a single location in the search domain with the best set of parameter values [Chu *et al.*, 2010] and estimate

parameters based on the Nash-Sutcliffe coefficient of efficiency (NSCE) as the objective function value.

Table 1. Parameter values of CREST model.

Parameter	Description	Unit	Min	Default	Max
RainFact	The multiplier in the precipitation field	----	0.5	1	2
Ksat	The Soil saturated hydraulic conductivity	mm/day	0	500	3000
MW	The Mean Water Capacity	mm	80	120	200
B	The exponent of the variable infiltration Curve	----	0.05	0.25	1.5
IM	The impervious area ratio	----	0	0.05	0.2
KE	The factor to convert the PET to local actual values	----	0.1	0.95	1.5
coeM	The overland runoff velocity coefficient	----	1	90	150
expM	The overland flow speed exponent	----	0.1	0.5	2

coeR	The multiplier used to convert overland flow speed to channel flow speed	----	1	2	3
coeS	The multiplier used to convert overland flow speed to interflow flow speed	----	0.001	0.3	1
KS	The overland Reservoir Discharge Parameter	----	0	0.6	1
KI	The interflow Reservoir Discharge Parameter	----	0	0.25	1

The initial condition of the CREST model is defined by three parameters of initial value of soil moisture, initial value of overland reservoir and initial value of interflow reservoir. The default values are taken as the initial values (Table 1). To let CREST be well distributed, a sufficient warm up period is necessary. For this study the warm up period of 6 months is given for the initial period of the calibration period of the CREST model.

2.9. Performance in Discharge Simulation

The performance of the model for the various precipitation products and global runoff products is evaluated using statistical comparisons between observed flow (O) and simulated flow (C) on different temporal resolutions. First, for statistical goodness of fit of simulated flows, we utilized the Nash-Sutcliffe coefficient of efficiency (Nash & Sutcliffe, 1970):

$$NSCE = 1 - \frac{\sum(Q_{i,o} - Q_{i,c})^2}{\sum(Q_{i,o} - \bar{Q}_o)^2} \quad (2)$$

Where $Q_{i,o}$ is the observed discharge of the i th day; $Q_{i,c}$ is the simulated discharge of i th day; and \bar{Q}_o the average of all the daily observed discharge values. If $NSCE \leq 0$, then the model provides no skill in relation to using the observed mean as a predictor and higher values indicating better agreement.

Second, the Pearson Correlation Coefficient (CC) is used to assess the agreement between simulated and observed discharge as follows:

$$CC = \frac{\sum(Q_{i,o} - \bar{Q}_o)(Q_{i,c} - \bar{Q}_c)}{\sqrt{\sum(Q_{i,o} - \bar{Q}_o)^2(Q_{i,c} - \bar{Q}_c)^2}} \quad (3)$$

Where \bar{Q}_c the average of all daily simulated discharge values.

Third, relative bias ratio assesses the systematic bias of the simulated discharge:

$$Bias = \frac{\sum Q_{i,c} - \sum Q_{i,o}}{\sum Q_{i,o}} * 100\% \quad (4)$$

Fourth, the Root Mean Square Difference (RMSD), the statistic most often used to quantify random differences between two variables, [Taylor, 2001] is defined by;

$$RMSD = \left[\frac{1}{N} \sum_{i=1}^N (Q_{i,o} - Q_{i,c})^2 \right]^{\frac{1}{2}} \quad (5)$$

Fifth, Standard Deviation (S), is a measure that is used to quantify the amount of variation or dispersion of a set of data values from its mean [Bland and Altman, 1996]. A low standard

deviation indicates that the data points tend to be close to the mean (also called the expected value) of the set, while a high standard deviation indicates that the data points are spread out over a wider range of values. The standard deviation shows the amount of variation of observed and simulated runoff from their respective mean value.

$$S = \sqrt{\frac{\sum_{i=1}^N (Q_{c,i} - \bar{Q}_c)^2}{N}} \quad (6)$$

Chapter Three

3. Evaluation of Satellite and Reanalysis Precipitation Products at Small scale:

A Case Study of Gilgel Abbay

3.1. Introduction

Water resource applications in regions without good quality and reliable temporal rainfall data are complicated by a lack of adequate spatial coverage. For example, in many African countries, the spatial coverage of hydrometeorological stations is very low. An African Climate Policy center (ACPC) assessment report [(ACPC), 2011] indicates while the spatial coverage of African climate stations is in the order of 1 station per 27,347 km², it is 1 station per 1244 km² in Germany. Similarly, the coverage of the streamflow stations in Africa is very low. There are 888 gauging stations in Africa in an area of 21,300,000 km² (without including the area of the Sahara), while there are 1150 streamflow gauges per 357,114 km² area in Germany. The situation in Ethiopia and upper Blue Nile is not different from this global picture. This contrasting revelation indicates the importance of supplementing the continent's hydro-information system with remotely sensed satellite and reanalysis products. Therefore, high-resolution precipitation products have major role in hydrological analysis and water resource application in such data-limited regions.

High-resolution satellite products have their own errors in reproducing streamflow hydrographs. It is well known that satellite rainfall values are estimates that are subject to a variety of error sources (e.g., gaps in revisit times, poor direct relationship between remotely sensed signals and rainfall rate, and atmospheric effects that modify the radiation field) and require a thorough validation [M M Bitew *et al.*, 2012; Dinku *et al.*, 2007; Hirpa *et al.*, 2010]. According to Dinku *et al.*, 2007; Hirpa *et al.*, 2010, validation efforts can be grouped into two categories. The first is the

direct comparison of the satellite rainfall estimates to the rain gauge networks and ground-based radar estimates. The second is the evaluation of satellite rainfall estimates based on their predictive ability of streamflow rate in a hydrological modeling framework. The second evaluation approach has two advantages. One, since the evaluation is performed at the watershed scale, it is not subject to the scale discrepancy problem that arises when using rain gauge data for validation. Two, the satellite rainfall estimates are evaluated with respect to a specific application, such as a driving input variable in a hydrologic model. Consequently, Dinku et al., 2007; Hirpa et al., 2010 evaluated each of the satellite products calibrated independently as input to hydrological models. M M Bitew et al., 2012 assessed four high resolution satellite precipitation products in the upper Blue Nile basin, Gilgel Abbay watershed: Climate Prediction Center's morphing technique (CMORPH), Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA) method near real-time (3B42RT) product (TMPA 3B42RT), TMPA method post-real-time research version (3B42) product (TMPA 3B42) and Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN); the study investigates the application of satellite precipitation products using the semi-distributed Soil and Water Assessment Tool (SWAT) hydrological model for a short period. They calibrated (2003–2004) and verified (2006–2007) the model independently with respect to each precipitation product. The result reveals that the 3B42RT and CMORPH simulations show consistent and modest performances in their simulations but underestimate the large flood peaks, while the 3B42 and PERSIANN simulations have inconsistent performance with poor or no skills. They also conclude the satellite-only product (3B42RT) performs much better than the satellite-gauge product (3B42), indicating that the algorithm used to incorporate rain gauge information with the goal of improving the accuracy of the satellite rainfall products is actually making the products worse, pointing to

problems in the algorithm. However, CMORPH and 3B42 satellite precipitation products show consistent and modest skills in their simulation in terms of volume and statistical performance for Gilgel Abbay watershed in the upper Blue Nile basin. Habib et al., 2014b also showed improvements of the bias-corrected CMORPH satellite data in the hydrological performance of Gilgel Abbay (1,656 km²) watershed compared to the uncorrected CMORPH product using the semi distributed Hydrologiska Byråns Vattenbalansavdelning (HBV) hydrological model for the years of 2003 and 2004. However, the uncorrected CMORPH shows modest performance based on the Nash-Sutcliffe coefficient of efficiency (NSCE) and relatively high bias values. These findings indicate many of the evaluations of the effects of satellite products on hydrological performance are tested by independently calibrating the models to each satellite product. Secondly, the different products indicated poor to modest skill for short calibration and validation periods at Gilgel Abbay watershed by applying semi distributed hydrological models.

This research aims to evaluate the performance of various satellite and reanalysis global precipitation datasets for water resource application in the upper Blue Nile basin, Ethiopia, using the Coupled Routing and Excess Storage (CREST) fully distributed hydrological model for multi-year and recent observed data from 2000 to 2011. Three precipitation products of gauge-adjusted CMORPH, (TRMM) TMPA 3B42v7 and the newly released ECMWF (Reanalysis) are used to validate the performance of the products in reproducing the flow hydrograph for three different calibration and simulation conditions (modes).

3.2. Data and Methods

3.2.1. Study Area

Gilgel Abbay watershed, which has a drainage area of 1,656 km², is located in Blue Nile Basin, in the Ethiopian part of the East Africa highlands between 36°48'E–37°24'E and 10°56'N–11°23'N

(Figure 12). The landscape has complex topography with elevation ranging between 1901–3336m, and the climate is semi-humid with mean annual rainfall of 1300mm. The land is covered by cropland (55%), forest and shrub (25%), and pasture (20%). Soils are dominantly fine textured: clay (42%), clay loam (39%), and sandy loam and silt (19%).

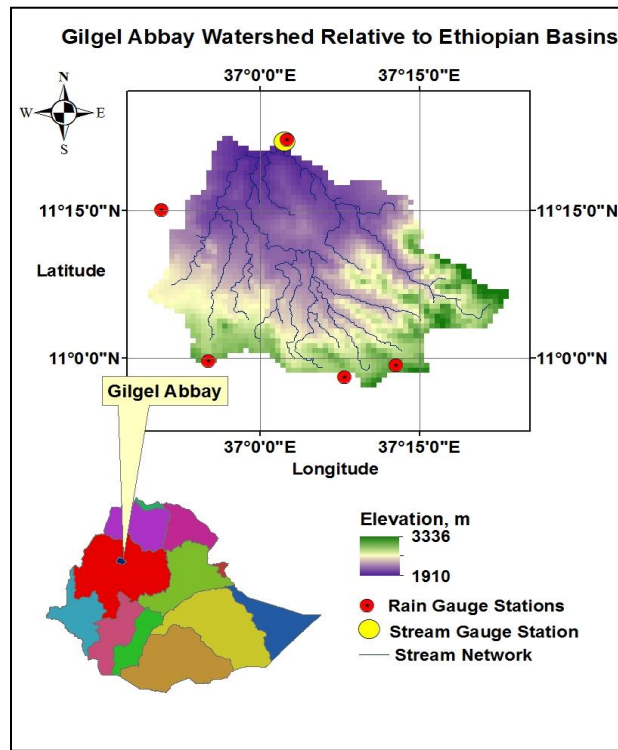


Figure 12. Gilgel Abbay watershed with locations of five nearby rain gauge stations of the study region, and one stream gauge station at the outlet of the watershed with 0.25° grid resolution.

3.2.2. Data Sets

In this study, three types of global precipitation datasets, with daily temporal scale (not aggregated) are taken as inputs: the high resolution ($0.25 \times 0.25^\circ$) gauge-adjusted (corrected) satellite precipitation datasets of CMORPH, TRMM (TMPA 3B42v7) and ($0.5 \times 0.5^\circ$) reanalysis forcing precipitation product of European Centre for Medium range Weather Forecasts (ECMWF). Station (gauge)-based daily measured rainfall and streamflow historical data (2000–2011) are taken as

inputs to setup the grid-based hydrological model (CREST). Grid-based daily temporal scale of Potential Evapotranspiration (PET, mm/day) data ($0.5 \times 0.5^\circ$) and reanalysis forcing precipitation ECMWF (daily, 0.5°) data are accessed from the earth2Observe Data portal (<https://wci.earth2observe.eu/portal/>). The grid-based (0.5°) reference potential evapotranspiration (PET) that is derived from the Penman-Monteith method is used as an input for the hydrologic simulations of Gilgel Abbay watershed.

The daily stream flow and the daily rainfall historical data are collected from the Ethiopia Ministry of Water Resources, Irrigation and Electricity and Ethiopian Meteorological Agency, respectively. The observed rainfall data is interpolated and gridded using Kriging method of interpolation with 0.25° grid resolution for the whole Abbay basin.

The visual inspection of the hydrograph shows that the baseflow record of Gilgel Abbay station o. showed an abrupt increase in late 2005. Habib et al., 2014b showed that the gauging station is moved by hundreds of meters downstream of the original site towards the end of 2005 due to road construction. This would have a great impact on the performance of the model for the late period of 2005 (validation period in our study). The Coupled Routing and Excess Storage (CREST) fully distributed hydrological model is setup for the evaluation of the precipitation products.

3.2.3. Parameters and Calibration

Many of the parameters in the CREST model can be estimated based on the availability of field survey data, such as soil surveys, land cover maps and vegetation coverage. There are approximately twelve parameters (Table 1) that are much more difficult to estimate from ancillary data and need to be calibrated either manually, automatically, or using combined approaches given observations of rainfall and streamflow.

Runoff simulations are carried out for various precipitation products that include in-situ gauge observations, gauge-adjusted CMORPH, (TRMM) TMPA 3B42v7 and ECMWF (Reanalysis). The model is run under three calibration and simulation conditions (modes). First, the model is calibrated and validated using in-situ gauge observed rainfall data and run for each precipitation forcing in a simulation mode (using rain gauge-calibrated parameters and hereafter named as ‘simulation mode’). Secondly, the model is calibrated and validated for each precipitation forcing independently (hereafter named ‘independent calibration mode’). Thirdly, precipitation adjustment is introduced to each precipitation forcing product through the parameter known as ‘RainFact’, which is the multiplier on the precipitation field and the adjustment factor of the precipitation either due to canopy interception or bias. Increasing the value of RainFact results in increasing runoff and vice versa and has a value greater than zero. The parameter RainFact is tuned for each product, and the model is run in simulation mode (hereafter named ‘RainFact calibration mode’). The analysis indicates the ‘RainFact’ parameter is highly sensitive to precipitation changes and helps adjust systematic bias embedded in the precipitation forcing. The model calibration is based on measured discharge data from 2000 to 2005, and validation is based on data from 2006 to 2011.

Table 2. Estimated parameter values of independent calibrations for different precipitation products.

Parameter	Description	Unit	Rain Gauge	CMORPH	TRMM (3B42v7)	ECMWF
RainFact	The multiplier in the precipitation field	----	1.399	1.3897	1.3994	1.8867

Ksat	The Soil saturated hydraulic conductivity	mm/day	143.3	117.87	116.63	35.212
MW	The Mean Water Capacity	mm	260.9	54.388	87.69	360.21
B	The exponent of the variable infiltration Curve	----	0.750	0.781	0.75525	0.15715
IM	The impervious area ratio	----	0.016	0.08906	0.016135	0.0555
KE	The factor to convert the PET to local actual values	----	1.193	0.82952	0.81581	0.8506
coeM	The overland runoff velocity coefficient	----	25.36	64.296	111.03	66.624
expM	The overland flow speed exponent	----	0.270	0.17103	0.297635	0.3109
coeR	The multiplier used to convert overland flow speed to channel flow speed	----	1.948	2.552	2.3768	1.5467

coeS	The multiplier used to convert overland flow speed to interflow flow speed	----	0.547	0.35733	0.8169	0.5881
KS	The overland Reservoir Discharge Parameter	----	0.107	0.08152	0.0693	0.0307
KI	The interflow Reservoir Discharge Parameter	----	0.081	0.06568	0.06782	0.0669

3.3. Result and Discussion

As described in the parameters and calibration section, the CREST model is run under three modeling conditions as simulation, independent calibration and RainFact calibration modes. For each product's independent calibration (Table 2), the examination of the model parameters revealed that the saturated hydraulic conductivity (Ksat), mean water holding capacity (WM) and the overland runoff velocity coefficient (coeM) show a wide range of parameter values for each product. Parameter Ksat varies from 35.21 to 143.36 mm/day, WM varies from 54.39 to 360.21 mm and coeM varies from 23.36 to 111.03 for the various precipitation products. These three parameters are less sensitive relative to the other parameters, and the model tries to pick any random value within the range of the parameters. It is also noted that the other sensitive model parameters remain relatively nearly the same for the tested precipitation products (Table 2). Among the twelve parameters, the most sensitive is RainFact, which has a high impact (from different independent parameters) on a given outcome.

CREST simulation performance was assessed using three commonly used statistical indices. For statistical goodness of fit of simulated flows, we utilized the NSCE. If $NSCE \leq 0$, then the model has no skill in relation to using the observed mean as a predictor. The Pearson Correlation Coefficient (CC) is used to assess the agreement between simulated and observed discharge. The relative bias ratio assesses the systematic bias of the simulated discharge. The best result occurs with $NSCE=1$, $CC=1$, and $Bias=0\%$

The performance of the model is assessed using gauge-based rainfall data to reproduce a hydrograph with NSCE of 0.82 for calibration and 0.67 for validation periods. The relative bias in both calibration and validation periods is within 5% of the observation for gauged rainfall (Table 3). The daily hydrograph of the three global precipitation products is well fitted for the independent calibration mode (Figure 13). Similarly, the results of the runoff hydrograph using the independent calibration mode shows a much better skill than previous work [*M M Bitew et al., 2012; Habib et al., 2014a*]. This may be attributed to the distributed nature of the CREST model. The CMORPH and ECMWF reproduce the runoff with NSCE of 0.78 and 0.75 for calibration and 0.71 and 0.72 for validation periods, respectively, under independent calibration mode. The performance of the TMPA 3B42v7 (TRMM) precipitation data is also encouraging with NSCE of 0.71 and 0.66 for the calibration and validation periods, respectively, under independent calibration (Table 3). All the global precipitation products underestimate the mean annual rainfall volume of the in-situ measured mean annual rainfall (Figure 14a, b). Generally, the mean annual runoff estimation decreases from CMORPH to TRMM for both calibration (2000–2005) and validation (2006–2011) periods shown in (Figure 14e, f) for independent calibration. TRMM underestimates the volume of flow relative to the other products with maximum bias of -16.3% and -36.6% in calibration and validation periods, respectively, for independent calibration (Table 3).

Table 3. Different calibration and simulation modes results of global precipitation products and gauged rainfall.

Precipitation Data	Independent Calibration Mode (Model Calibrated to Each Precipitation Input)	Simulation Mode (Using Calibrated Gauged Rainfall Parameters)	RainFact Calibration Mode (Using Calibrated Gauged Rainfall Parameters and Only RainFact Parameter Is Calibrated)
Gauged Rainfall	Calibration 2000–2005		
	NSCE = 0.82	NSCE = 0.82	NSCE = 0.82
	Bias(%) = -4.6	Bias(%) = -4.6	Bias(%) = -4.6
	CC = 0.91	CC = 0.91	CC = 0.91
	Validation 2006–2011		
	NSCE = 0.67	NSCE = 0.67	NSCE = 0.67
	Bias(%) = -5	Bias(%) = -5	Bias(%) = -5
	CC = 0.89	CC = 0.89	CC = 0.89
CMORPH	Calibration 2000–2005		
	NSCE = 0.78	NSCE = 0.72	NSCE = 0.75
	Bias(%) = -6.5	Bias(%) = -21.9	Bias(%) = -4.8
	CC = 0.88541	CC = 0.86497	CC = 0.87

	Validation 2006–2011		
	NSCE = 0.71	NSCE = 0.60	NSCE = 0.68
	Bias(%) = -25.9	Bias(%) = -39.9	Bias(%) = -26.0
	CC = 0.88	CC = 0.87	CC = 0.87
TRMM	Calibration 2000–2005		
	NSCE = 0.71	NSCE = 0.65	NSCE = 0.70
	Bias(%) = -16.3	Bias(%) = -31.5	Bias(%) = -19.2
	CC = 0.87	CC = 0.86	CC = 0.86
	Validation 2006–2011		
	NSCE = 0.66	NSCE = 0.50	NSCE = 0.64
	Bias(%) = -36.6	Bias(%) = -50.0	Bias(%) = -38.0
	CC = 0.89	CC = 0.87	CC = 0.88
	ECMWF	Calibration 2000–2005	
NSCE = 0.75		NSCE = 0.57	NSCE = 0.73
Bias(%) = -1.6		Bias(%) = -42.0	Bias(%) = -6.5
CC = 0.87		CC = 0.85	CC = 0.86
Validation 2006–2011			

	NSCE = 0.72	NSCE = 0.49	NSCE = 0.68
	Bias(%) = -12.9	Bias(%) = -49.3	Bias(%) = -17.6
	CC = 0.87	CC = 0.86	CC = 0.86

For the simulation mode (using calibrated gauged rainfall parameters), the performance in capturing the mean annual runoff is also encouraging for TRMM and CMORPH as shown in Figure 13c, d. The statistical performance (Table 3) shows CMORPH and TRMM perform relatively well with respective NSCE values of the calibration period 0.72 and 0.65, as well as 0.6 and 0.5 for the validation period for the simulation mode. The performance of the reanalysis ECMWF satellite product for the simulation mode shows a relatively lower NSCE value of 0.57 for calibration and 0.49 for validation.

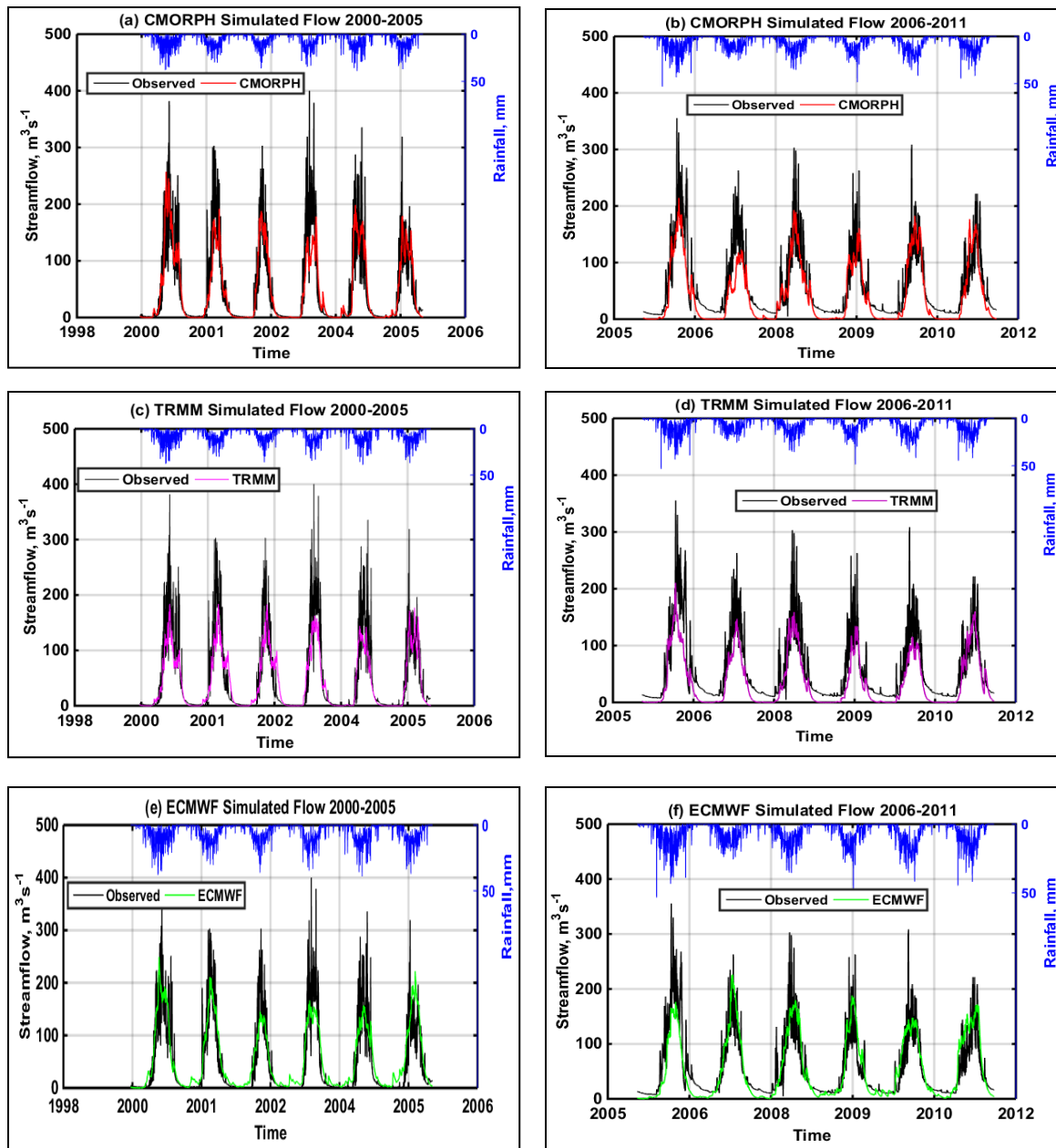
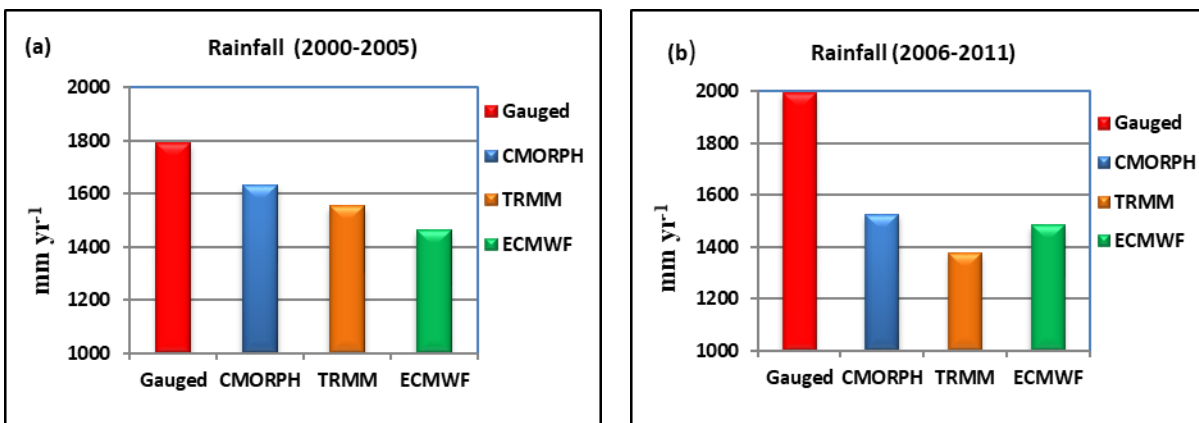


Figure 13. Hydrograph of simulated flow of global precipitation products with observed flow of Gilgel Abbay watershed for the Independent calibration mode (a,b) CMORPH, (c,d) TRMM, (e,f) ECMWF.

The performance of all precipitation products for the independent calibration mode indicates the global precipitation products capture the seasonal patterns of the runoff hydrograph (Figure 13). Further work is required to improve the annual volume of the precipitation products, especially the reanalysis product (ECMWF) as shown in Figure 14a, b. As a way of improving the performance in a simulation mode, we fine-tune the rainfall factor (RainFact), which is one of the parameters in the CREST model and has a major role in multiplying the precipitation products in the field ranges from 0.5 to 2 to decrease the bias in the precipitation input for each precipitation product, and the results are improved significantly (Figure 14g, h). Figure 14 shows that the inherent bias in the mean annual volume of the simulated flow of each precipitation product emanates from the underestimation of the in-situ gauge-observed rainfall of the global precipitation products. The validation period shows consistently lower performance than the calibration period for all precipitation products. The low performance of the validation period may be related to the change of the location of the streamflow gauging station [Habib *et al.*, 2014a]. The station is moved hundreds of meters downstream of the original site towards the end of 2005 due to road construction.



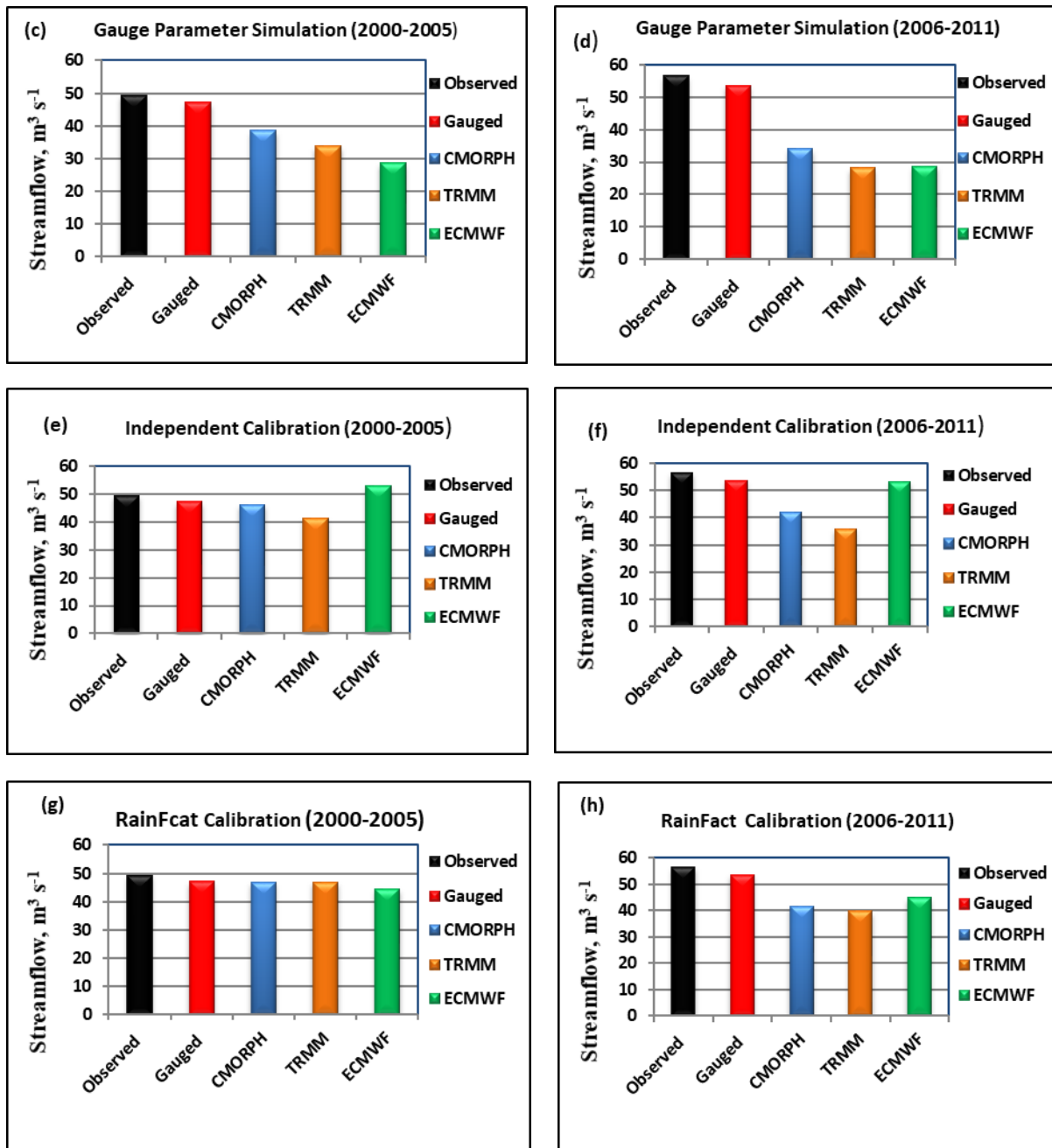


Figure 14. Different precipitation products comparison of (a,b) Average annual rainfall, (c,d) Average annual flow for Simulation mode, (e,f) Average annual flow for Independent calibration mode, (g,h) Average annual flow for RainFact calibration mode.

Evaluation of the flow quantiles using Quantile-Quantile (Q-Q) plots is shown in Figure 15. The Quantile-Quantile (Q-Q) plot is a graphical technique to determine if two datasets come from populations with a common distribution. A 45-degree reference line is also plotted. If the two sets come from a population with the same distribution, the points should fall approximately along this reference line. The greater the departure from this reference line, the greater the evidence for the conclusion that the two datasets come from populations with different distributions or with the same distribution but biased or shifted with respect to the mean. The Q-Q plot of the quantiles from the simulation mode (Figure 15b) shows all the precipitation products have high departure from the reference line except the in-situ rain gauge-simulated quantiles. The simulated flows of the global precipitation datasets consistently underestimate the quantiles in simulation modes (Figure 15b). In Q-Q plots (Figure 15a, c), the flow quantiles plots cluster very well around the reference line. Particularly, the RainFact mode plot captures the higher quantile values and is better than other independent calibration and simulation modes. Overall, the result reveals that tuning the parameter 'RainFact', which is a rainfall multiplier, shows to perform better than independent calibration (of all parameters) in Q-Q plot evaluation to capture the peak flows for all the precipitation products (Figure 15c).

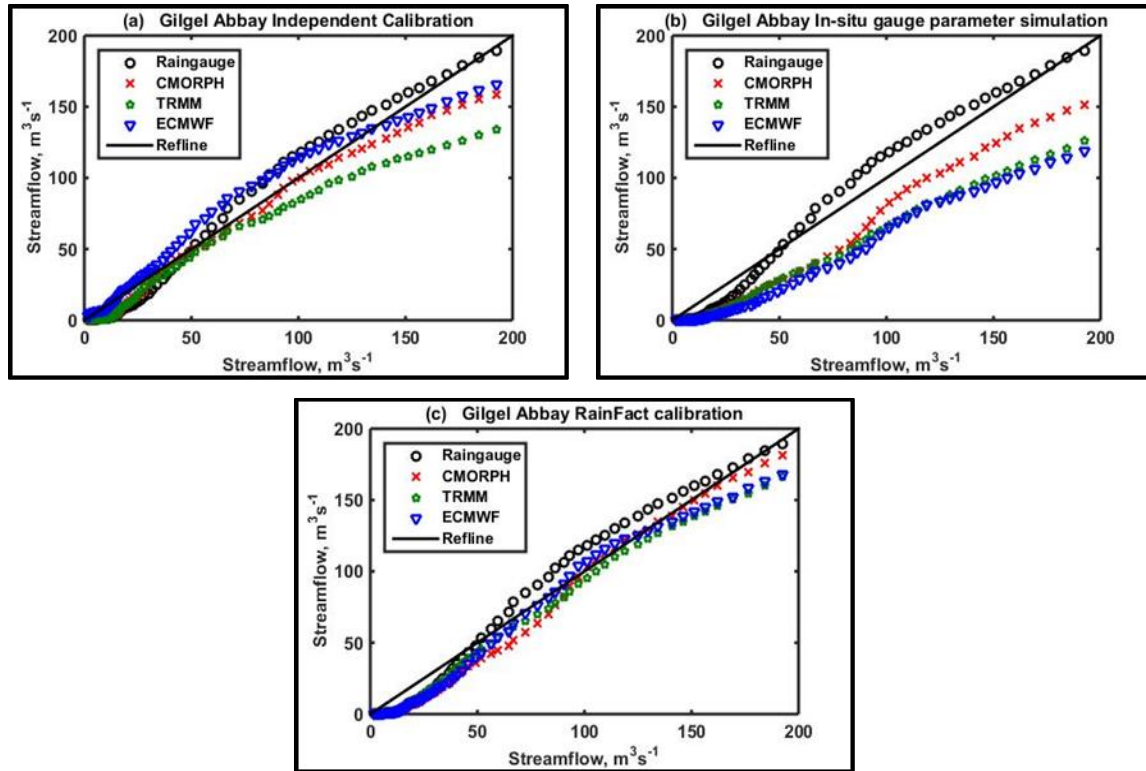


Figure 15. Quantile-Quantile plot of Gilgel Abbay watershed for (a) Independent calibration, (b) In-situ gauge parameter simulation, (c) RainFact calibration.

3.4. Conclusive Remarks

The aim of this study is to assess the performance of various satellite and reanalysis precipitation datasets for water resource applications in the upper Blue Nile basin, Ethiopia, applying different calibration and simulation conditions (modes). The Coupled Routing and Excess Storage (CREST V.2.1) fully distributed hydrological model is setup for different simulation modes for Gilgel Abbay watershed (Area = 1,656 km²) in the upper Blue Nile basin.

Examination of the parameters of the model indicates Saturated Hydraulic Conductivity (Ksat), Mean Water holding capacity (WM) and the overland runoff velocity coefficient (coeM) show wide range parameter values for the different precipitation products of independent calibration. These three parameters are less sensitive relative to the other parameters. It is also noted that other

model parameters are sensitive and remain nearly the same for the tested precipitation products. The most sensitive parameter is RainFact.

The in-situ gauge observed rainfall data reproduces the runoff very well (NSCE = 0.82) as compared to the observed daily discharge at the outlet of the watershed at Gilgel Abbay watershed, indicating the CREST model has the ability to capture the hydrological processes and to evaluate the global precipitation products.

In terms of comparing the performance of the satellite and reanalysis precipitation products, all products under independent and RainFact calibration modes reproduce daily streamflow better than the simulation mode with NSCE performance above 70% in calibration and 60% invalidation periods.

In simulation mode, though the NSCE performance is reasonable (CMORPH and TRMM), the runoff volume and extreme flow ranges are underestimated as indicated by the bias and Q-Q plots. The ECMWF reanalysis product underestimates the runoff results more than other tested precipitation products. However, the ECMWF reanalysis product has shown significant improvement in independent calibration and RainFact calibration modes to capture the reference line on the simulated high flows. Generally, the RainFact calibration mode enhances the simulated flows in capturing high flows of the observed streamflow in the Q-Q plot.

The validation period shows consistently poorer performance than the calibration period for all precipitation products due to the movement of the streamflow gauging station, i.e., hundreds of meters downstream of the original site towards the end of 2005 due to road construction. Therefore, we can conclude that satellite and reanalysis precipitation products can be used for water resource planning and management under independent calibration mode. Further analysis is underway to

understand the response characteristics of the precipitation products at a larger scale of Blue Nile basin.

The results from the different calibration modes show that the independent calibration is time consuming and demands several times for calibration and it is exclusively applicable for small spatial scale watersheds applying fully distributed hydrological model. For larger basin scales in the upper Blue Nile basin, considering various precipitation products and carrying out independent calibration of specific product using fully distributed hydrological model CREST is much time taking. For this understanding, it is recommended to carry out RainFact calibration mode for each precipitation product which presents an equivalent performance with the independent calibration and improves the performance much from the simulation mode. The next chapter considers different sized watersheds (small, medium and large basin scales) and several precipitation products, and the evaluation of the products has made applying the RainFact calibration and simulation modes.

Chapter Four

4. Multi-Scale Evaluation of Globally Available Satellite and Reanalysis Precipitation Products in the Upper Blue Nile Basin (Abbay), Ethiopia.

4.1. Introduction

Hydrometeorological datasets have significant importance on the sustainable water resources management, as well as for different socio-economic activities. Evaluation of the amount of different reliable precipitation products and of its spatiotemporal distribution has significant role for scientific hydrological applications including but not limited to increasing our understanding of the hydrological cycle, assessing the hydrological impacts of human activities, assessing water resources, irrigation planning, and forecasting of droughts and floods [*H Beck et al.*, 2017; *Komma et al.*, 2007; *Kucera et al.*, 2013; *Lettenmaier et al.*, 2015].

It is noted that acquiring accurate and consistent set of hydrometeorological data is a challenging task throughout the world especially in Africa. The data assessment report of Africa-African Climate Policy Centre [(*ACPC*), 2011] summarized as having poor spatial coverage with many stations became either non-operational or report poor quality data with missing information. It is indicated that in some cases even the existing stations neither provide accurate observations nor quality controlled and transmitted to the database system. Hence, the importance of improved global data remains one of the future hopes for the design and management of water resources in data scarce parts of the world.

The advent of global datasets such as satellite and reanalysis products play a great role in building a consistent set of hydrometeorological data that can be utilized for various purposes of water resources development and management in Africa and globally. Improved globally available

products are the future hope to many regions of the world where the quantity and quality of data is deteriorating [(ACPC), 2011]. Numerous hydrological performance researches are conducted globally on various global satellite precipitation products [M Bitew and Gebremichael, 2011; Habib et al., 2014b; Mei et al., 2014; Seyyedi et al., 2015a]. M Bitew and Gebremichael, 2011 evaluated the hydrological performance of CMORPH, satellite-only product TRMM TMPA 3B42RT, Satellite-gauge product TRMM TMPA 3B42 and Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) for small size watersheds of Gilgel Abbay (1,656 km²) and Koga (299 km²) in Abbay basin, Ethiopia. Applying a semi distributed hydrological model SWAT calibrated from 2003 to 2004 and validated from 2006 to 2007. Granting to the subject area, the performance of the results of the simulations obtained from the 3B42RT and CMORPH products improves with the size of watershed area increases. Habib et al., 2014a also showed the improvements of the bias corrected CMORPH satellite data on the hydrological performance of Gilgel Abbay (1,656 km²) watershed compared to the uncorrected CMORPH product. The products are simulated using semi distributed Hydrologiska Byråns Vattenbalansavdelning (HBV) hydrological model from 2003 to 2004. The above studies that are conducted in Abbay basin, Ethiopia, and carried out in small watersheds and for short simulation periods. In most of these studies, all precipitation products are calibrated independently (specific product calibration) and compared.

Similar studies are conducted through the world to adapt to and utilize the global datasets for various purposes [H Beck et al., 2017; Mei et al., 2014; Seyyedi et al., 2015a]. For instance, [Mei et al., 2014] investigated error characteristics of six quasi-global satellite precipitation products and their error propagation in flow simulations for a range of mountainous basin scales (255–6,967 km²) in the upper Adige river basin (Eastern Italian Alps) for different periods (May–August and

September–November). The results suggest a positive correlation between the systematic error and basin elevation providing clues for further improvement of the products. Performance evaluation of flow simulations yields a higher degree of consistency for the moderate to large basin scales and the May–August period. Gauge adjustment for the different satellite products is shown to moderate their error magnitude and increase their correlation with reference precipitation and streamflow simulations. Seyyedi et al., 2015a also assessed the error propagation of two global (or near-global) precipitation datasets in terms of flood modelling for a range of basin scales (300–70,000 km²) focusing on multi-year (2002–2011) simulations over a mid-latitude basin (Susquehanna River Basin) in the North-eastern United States. The datasets are the TRMM 3B42v7 research product, and the Global Land Data Assimilation (GLDAS) reanalysis system precipitation dataset. Results show that the satellite product exhibits significantly better error statistics compared to the GLDAS. Particularly for the simulated streamflow, GLDAS is shown to have up to 7(3) times higher mean relative error compared to the corresponding TMPA 3B42v7 error metric for moderate (extreme) streamflow values. H Beck et al., 2017 evaluated the performance of Multi-SourceWeighted Ensemble Precipitation (MSWEP) version 1.1 using simple conceptual hydrological model HBV for 9011 catchments across the globe. For the 1058 sparsely gauged catchments, representative of 83.9% of the global land surface (excluding Antarctica), MSWEP obtained a median calibration NSCE of 0.52 vs. 0.29–0.39 for the other Precipitation datasets.

This research mainly aims to evaluate and compares the global satellite and reanalysis precipitation products for their hydrological performance and application potential for water resources applications in data scarce basins. The study is conducted in the upper Blue Nile basin over three different sized watersheds (1,656–199,812 km²) and multi-year (2000–2012) using a daily

temporal scale. The study focuses on the hydrological performance of two satellite and three reanalysis global gridded precipitation products (0.25°) resolution in simulating the daily flow time series over the selected watersheds. The two satellite datasets included are the gauge adjusted CMORPH, and TRMM TMPA 3B42v7 and the three reanalysis products are the ERAI, GPCC and MSWEP. The Coupled Routing and Excess Storage (CREST) fully distributed hydrological model [Xinyi Shen and Hong, 2014] is used for this evaluation and comparison study. Unlike many studies shown above which compares the results of independently calibrated performance, this study approach follows comparison of the simulation results of different products using one set of model parameters that is calibrated with in-situ gauge observed rainfall data. It means each time we run the model for each precipitation product, we maintain the gauge calibrated model parameter sets. Secondly, we undertake sensitivity analysis and tweak the dominant parameter of the model to each product and reevaluate the performance of each product. This approach to a certain extent may indicate whether the sources of the bias is from the input data or the model structure.

4.2. Data and Methods

4.2.1. Study Area

Three nested watersheds (small, medium and large) in the upper Blue Nile basin that represent different spatial scales are considered in this study: Gilgel Abbay (small-scale) which is located upstream of Lake Tana, Abbay at Kessie (medium-scale) and Abbay at border or Eldiem station (large-scale) (Figure 16). Brief description of the watersheds is provided below.

Small scale (Gilgel Abbay Watershed)

Gilgel Abbay is a small watershed located at the upper part of Abbay basin between 36°48'E–37°24'E and 10°56'N–11°23'N. It is the major tributary of Lake Tana and has a drainage area of 1,656km². The landscape is complex topography with elevation range between 1,901-3,336m. The mean annual runoff of Gilgel Abbay varies from 41 to 60 m³s⁻¹ and the seasonal mean seasonal runoff in the wet season of the four months (JJAS) varies between 71 to 115 m³s⁻¹ based on the data length used in this study (2000 to 2012).

Medium Scale (Kessie Station)

Kessie gauging station is located at the mid of Abbay basin (including Gilgel Abbay watershed and Lake Tana), which has a drainage area of 65,784 km², in the upper Blue Nile basin between 36°42'E-39°48'E and 9°15'N-12°45'N. The land scale represents complex topography with elevation range between 1,102-4,176m. The mean annual runoff of Kessie varies from 450 to 950 m³s⁻¹ and seasonal mean runoff in the wet season of the four months (JJAS) of Abbay at Kessie varies between 870 and 1,634 m³s⁻¹. It is noted that the observed flows at Kessie are affected by water released from Lake Tana, which represent about 20% of the monthly flows at Kessie.

Large scale (Eldiem)

Eldiem station is located at the outlet of Abbay basin (boarder of Ethiopia and Sudan) having a drainage area of 199,812 km² (including Gilgel Abbay and Kessie) that represents the main part of Ethiopian highlands between 34°24'E–39°48'E and 7°42'N–12°30'N. The landscape is complex consisting of plain (flatter) area on the North-West part and highly elevated mountains on the other parts of the basin with elevation ranging between 503 and 4,176m. The mean annual runoff of Eldiem varies between 1,249 and 1,981 m³s⁻¹ and the seasonal mean runoff in the wet season of the four months (JJAS) varies between 2,412 and 3,344 m³s⁻¹. The Grand Ethiopian Renaissance

Dam (GERD), formerly known as the Millennium Dam and sometimes referred to as Hidase Dam, is a gravity dam on the Blue Nile River in Ethiopia that has been under construction since 2011. The dam is situated close to the border of Ethiopia and Sudan near to at the upstream of the Eldiem streamflow gauging station. The evaluation of the global products with respect to Eldiem represents the GRED dam correspondingly, since the station is located too close to the Eldiem station. The daily historical streamflow data of the border is not available in the Ethiopian water, irrigation and electricity and the daily data is collected from Sudanese government.

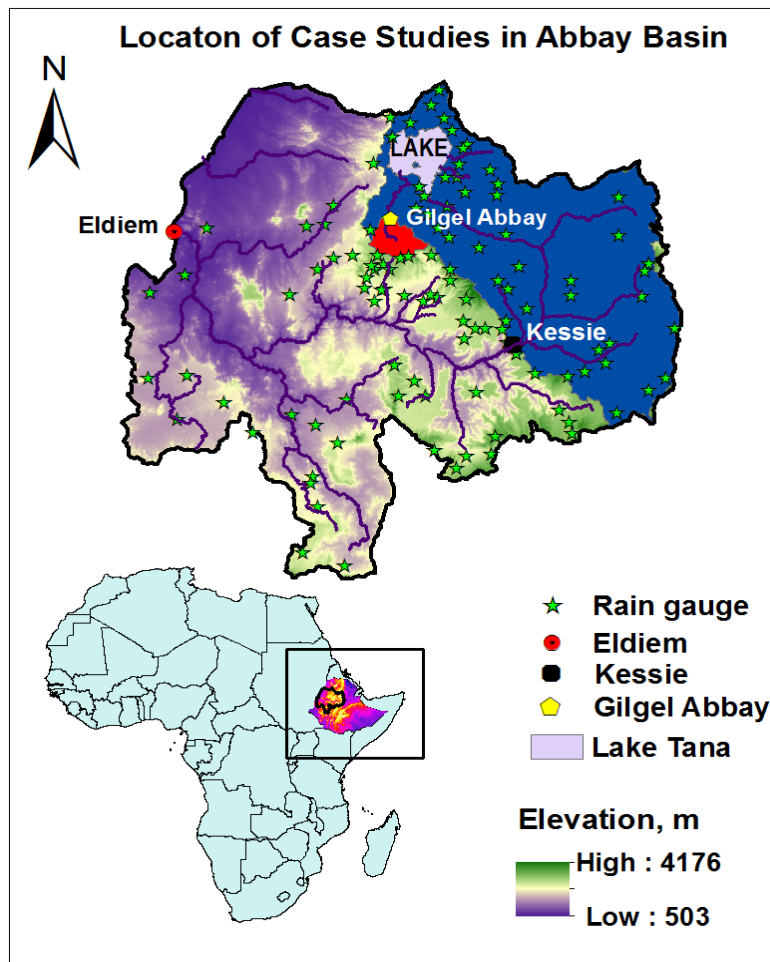


Figure 16. Location of the three nested watersheds and rain gauge stations distribution in the upper Blue Nile basin.

4.2.2. Data Sets

The daily stream flow of the three case studies and the daily rainfall historical data are collected from Ethiopia Ministry of Water, Irrigation and Electricity, and Ethiopian Meteorological Agency respectively. A total of 89 in-situ rainfall measuring stations are used for the three case studies. The observed rainfall data is interpolated using Kriging method using the available 89 stations with 0.25o spatial resolution by Matlab. The grid based (0.5°) with daily temporal scale of reference potential Evapotranspiration (PET, mm/day) that is derived from Penman-Monteith method [Richard G. Allen *et al.*, 1998] has been used as an input to setup CREST model for the hydrologic simulations. The grid PET data is accessed from the earth2Observe data portal (<https://wci.earth2observe.eu/portal/>). The Coupled Routing and Excess Storage (CREST) fully distributed hydrological model is set up for the evaluation of the precipitation products.

In this study satellite and reanalysis precipitation products with daily temporal scale are considered for hydrological evaluation of daily simulation flow for different basin scale in the upper Blue Nile basin. Satellite precipitation products of gauge-adjusted CMORPH and TRMM TMPA 3B42v7 and reanalysis precipitation products of ERAI, GPCC and MSWEP with a resolution of (Daily, 0.25°) are considered in the hydrological evaluation of this study.

4.2.3. Parameters and Calibration

The CREST model has 12 parameters. In theory all these parameters can be derived from field survey data, such as soil, land and vegetation cover information and calibration may not be required. However, no physically based distributed hydrological model is immune from calibration when applied to in the context of big basins for obvious reasons of practical challenges of acquiring these parameter values at each grid location.

In this study, two step approaches are used to calibrate the CREST model. Manual calibration of multi-parameter model is time consuming and less practical. First, automatic calibration has been implemented to estimate the model parameters. CREST uses, Shuffle Complex Evolution University of Arizona (SCE-UA) [Duan *et al.*, 1992] automatic kernel algorithm for calibration of the model using in-situ gauged rainfall. CREST directly uses the screen output of SCE-UA codes in Matlab written by [Duan *et al.*, 1992]. The objective function value is the NSCE of each simulation. Second, sensitivity analysis is done to identify the most sensitive (or dominant) parameter. The automatic calibration results are shown in Table 4. In our previous paper [Lakew *et al.*, 2017], the RainFact parameter, which is the multiplier on the precipitation field, is identified as the most dominant parameter controlling the accuracy of the model performance. This parameter is tweaked and performance improvements evaluated for each watershed and precipitation products.

The CREST model is calibrated for the three case studies using daily in-situ gauge observed rainfall, potential evapotranspiration (PET), and observed streamflow data. Calibration is done from 2000 to 2007 and validated from 2008 to 2012.

Table 4. Estimated parameter values of independent calibration for different case studies.

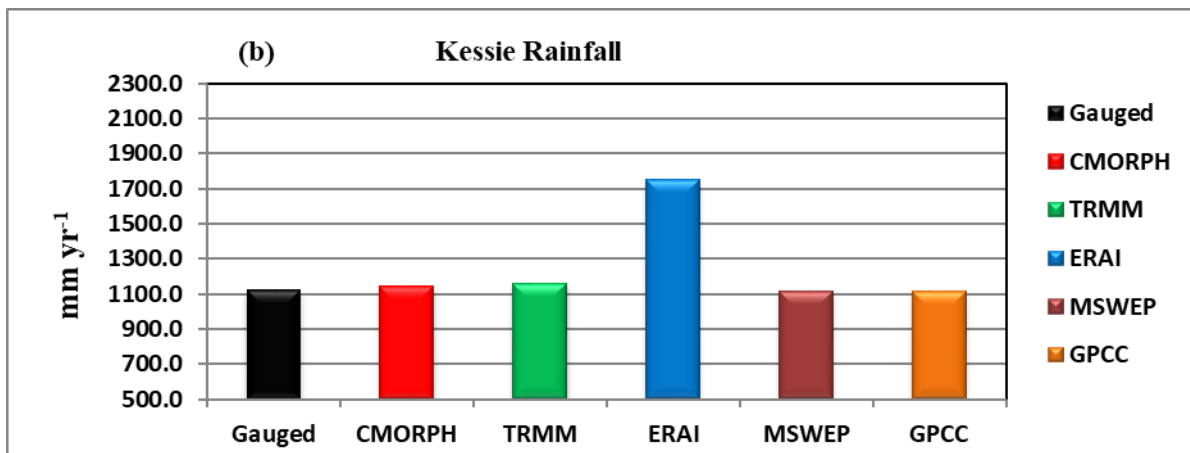
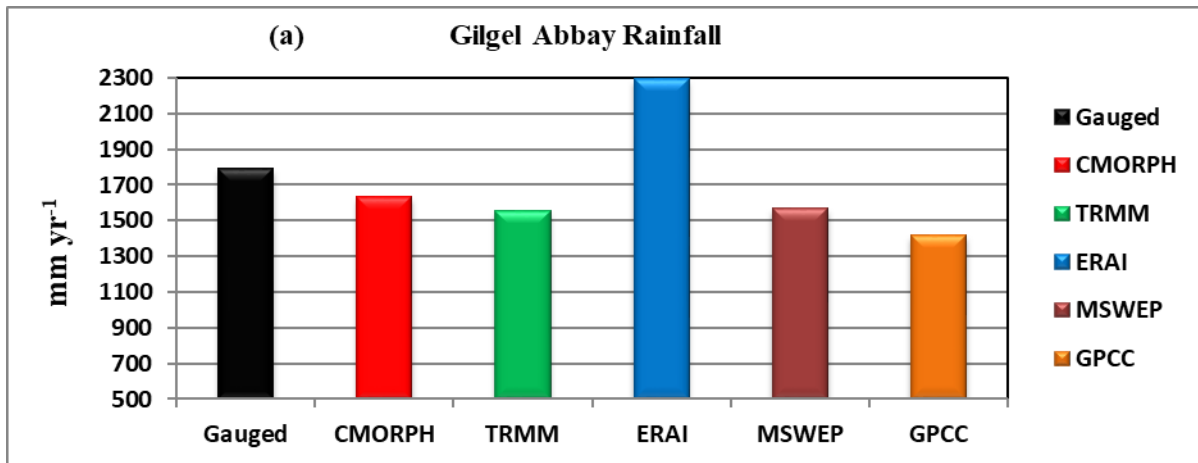
Parameter	Description	Unit	Multi-Scale case studies		
			Gilgel Abbay	Kessie	Eldiem
RainFact	The multiplier on the precipitation field	----	1.4	0.76	0.67

Ksat	The Soil saturate hydraulic conductivity	mm/day	143.3	177.7	151.5
MW	The Mean Water Capacity	mm	260.9	470.6	360.6
B	The exponent of the variable infiltration Curve	----	0.75	1.09	0.87
IM	The impervious area ratio	----	0.016	0.167	0.067
KE	The factor to convert the PET to local actual	----	1.19	0.82	1.02
coeM	The overland runoff velocity coefficient	----	25.36	135.75	6.63
expM	The overland flow speed exponent	----	0.270	0.46	0.11
coeR	The multiplier used to convert overland flow speed to channel flow speed	----	1.94	2.93	2.12
coeS	The multiplier used to convert overland flow speed to interflow flow speed	----	0.54	0.029	0.74
KS	The overland reservoir Discharge Parameter	----	0.31	0.37	0.58
KI	The interflow Reservoir Discharge Parameter	----	0.18	0.27	0.17

4.3. Result and Discussion

Rainfall Analysis

The areal average global precipitation products used in this study are compared with the in-situ gauged rainfall data (Figure 17). The areal mean annual observed rainfall of Gilgel Abbay, Kessie and Eldiem case studies for the study period of 2000 to 2012 is 1750, 1100 and 1400 mm yr^{-1} respectively. In comparison to the in-situ gauged rainfall, most of the global precipitation products exhibit good agreement with gauged data except the ERAI product which overestimates in all the three case studies. It is also noted that the accuracy of the areal average precipitation slightly increases as a function of the size of the watershed.



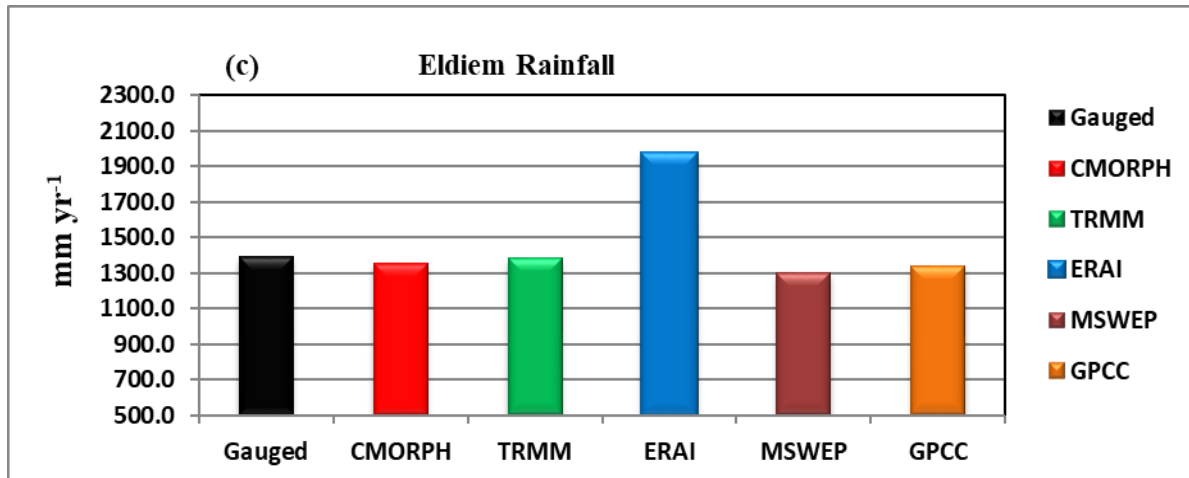


Figure 17. Mean annual rainfall of (a) Gilgel Abbay (b) Kessie and (c) Eldiem case studies.

Runoff Analysis

Runoff simulations are carried out for the two satellite precipitation products of gauge-adjusted CMORPH and (TRMM) TMPA 3B42v7 and the three reanalysis precipitation products of ERAI, MSWEP and GPCC described in global gridded precipitation products section. The model is calibrated using in-situ gauge observed daily rainfall data for 8 years period (2000-2007) and validated for 5 years (2008-2012). The final list of parameters is shown in Table 4.

For illustration the focus is on the RainFact parameter (sensitive) which is the multiplier on the precipitation field and the adjustment factor of the precipitation either due to canopy interception or bias. Increasing the value of the RainFact result in increasing runoff and vice versa and has value of greater than zero. If the parameter value is greater than one, it shows that the average areal rainfall amount that is taken for that watershed underestimate the actual average areal rainfall of the region to capture the observed flow and vice versa. The effect of the parameter shows that as the scale of the watershed increases the RainFact parameter value decreases from 1.4 to 0.67. As

the basin scale increases, it has the tendency to have more lowlands and, in this study, more gauging stations are taken from highlands that have a high amount of rain. This makes, as the basin scale increases the tendency to overestimate the actual rainfall of the region increases. To correct this effect CREST uses the RainFact value less than one of 0.67 for Eldiem (large scale) and 0.76 for Kessie (medium scale) that have relatively large and medium lowland coverage respectively. Gilgel Abbay watershed has well distributed rain-gauges that are set up in the central highlands of the basin, the RainFact parameter is expected to be close to 1. However, the parameter is 1.4 which is more than the expected value. This indicates that there is a high groundwater contribution in the streamflow of Gilgel Abbay watershed that makes the RainFact parameter value greater than 1. Melesse and Abtew, 2015 in the study of Landscape Dynamics, Soils and Hydrological process in Varied Climate they strengthen this idea and said that the hydrological process governing runoff generation in Gilgel Abbay is dominated by saturation excess and base flow contribution.

The performance of the various precipitation products for Gilgel Abbay watershed applying the RainFact calibration (R) and simulation (S) modes for the calibration period (Cal) and validation period (Val) is shown in Table 5. Gilgel Abbay watershed that has relatively very good rain gauge coverage, the simulated flows from different products applying the simulation mode (S) perform very well (except ERAI) in the NSCE evaluation (from 0.65 to 0.76) at calibration period. Especially MSWEP shows better performance in small scale of the basin (NSCE =0.76), and CMORPH also shows good performance of NSCE=0.72 under the simulation mode (S). The reanalysis product of ERAI shows relatively modest skill of NSCE = 0.55. The reanalysis product of GPCP (0.65) shows encouraging performance with over the satellite products of CMORPH (0.72) and TRMM (0.64) in the calibration period under simulation mode (S). This shows that the newly released reanalysis products (except MSWEP) don't show significant improvement over the

satellite products to capture the total volume of the runoff in Gilgel Abbay watershed under the simulation mode (S). Considering this, we introduce the rainfall adjustment factor, RainFact calibration (R) mode for each precipitation product to improve the simulation efficiency of each product.

The RainFact calibration mode of Gilgel Abbay shows that each product shows significant improvement of NSCE value. Especially the MSWEP product simulation flow applying RainFact calibration mode shows the highest improvement of NSCE=0.81 and 0.70 for calibration and validation periods respectively from both satellite and reanalysis precipitation products as shown in Table 5.

Based on the bias (%) performance evaluation in the Gilgel Abbay watershed, the simulated flows from both satellite and reanalysis products underestimate the observed flow and relatively CMORPH shows the minimum bias value of (-21.8%) in the simulation mode (S). The RainFact calibration mode (R) bias values shows that the simulated flow from CMORPH shows the minimum bias value of (-4.8%) and shows high improvement in the calibration period. This indicates that the simulated flow of satellite precipitation product of CMORPH outperforms the reanalysis products based on bias performance in the Gilgel Abbay watershed.

The Correlation Coefficient (CC) evaluation shows that each product shows very good coefficient performance of above 0.84 for both calibration and validation periods. Especially, MSWEP shows the highest of 0.92 in the calibration period under RainFact calibration (R) mode. These all show that based on statistical metric performance measurement, applying RainFact calibration mode for MSWEP product shows high performance and improvement over the rest precipitation products, and it can be an alternative data source for water resources applications in the Gilgel Abbay watershed, if rainfall adjustment factor is introduced.

Table 5. Daily performance of various precipitation products of the Gilgel Abbay watershed.

		Precipitation Products										
		Gauge Calibrated	CMORPH		TRMM		ERA-Interim		MSWEP		GPCP	
			S	R	S	R	S	R	S	R	S	R
NSCE	Cal	0.82	0.72	0.75	0.65	0.70	0.55	0.74	0.76	0.81	0.65	0.80
	Val	0.67	0.60	0.68	0.5	0.64	0.19	0.58	0.64	0.70	0.43	0.61
Bias	Cal	-4.6	-21.9	-4.8	-31.6	-19.3	24.4	-24.9	-29.4	-18.9	-40.5	-18.6
	Val	-5	-39.9	-26	-50	-38	28.7	-34.4	-38.6	-28	-54.1	-36.4
CC	Cal	0.91	0.87	0.87	0.86	0.86	0.87	0.88	0.91	0.92	0.90	0.90
	Val	0.89	0.87	0.87	0.87	0.88	0.84	0.84	0.88	0.89	0.85	0.86

Where, S= Simulation mode, R= RainFact calibration mode, Cal= Calibration period (2000-2007), Val= Validation period (2008-2012).

The medium spatial scale of this study (Kessie) has the effect of Lake Tana, which is located at the upstream of the streamflow gauging station of Kessie. The daily performance of NSCE for the simulated flow from a locally set hydrological model of CREST using gauged rainfall and the simulated flows for the whole products shows less performance of NSCE less than 0.72 in Kessie relative to the Gilgel Abbay watershed. This is imputable to the effect of Lake Tana that has a reservoir surface area ~3,500 square km located at the upstream of Kessie streamflow gauging station.

For medium scale basin (Kessie, to the effect of Lake Tana) the simulated flow of reanalysis products shows the better NSCE performance of MSWEP (0.71) and GPCP (0.69) over the satellite products of CMORPH (0.61) and TRMM (0.66) in the calibration period under the

simulation mode (S) as shown in Table 6. The newly introduced RainFact calibration mode (R) doesn't show any significant improvement over the simulation mode (S) contrary with the Gilgel Abbay watershed. This indicates that the simulated flow of the whole precipitation products has less performance in Kessie that has the lake effect on both simulation modes. However, the product from MSWEP shows better NSCE evaluation performance of 0.72 and 0.71 for RainFact calibration (R) and simulation (S) modes respectively in the calibration period. The product from ERAI shows the worst skill performance of NSCE with a negative value for both RainFact calibration and simulation modes.

The bias performance of Kessie shows that the same with Gilgel Abbay watershed, the satellite driven simulation flow from CMORPH manifests the minimum bias value of -10.9% under the simulation mode in calibration period. However, the bias value under RainFact calibration mode (R) doesn't show any improvement for the CMORPH simulation (S) mode for both calibration and validation periods.

The medium scale of Kessie station has the same Correlation Coefficient (CC) efficiency with the Gilgel Abbay with very good performance of above 0.76 for the whole product under both simulation modes for both calibration and validation periods. The product from MSWEP shows the best correlation coefficient of 0.87 and 0.80 for calibration and validation periods under RainFact calibration modes. The newly introduced RainFact calibration mode shows modest improvement over the simulation mode (S) for both satellite and reanalysis precipitation products for the Kessie station that has a reservoir effect of Lake Tana. The product from MSWEP under the newly introduced RainFact calibration mode shows better performance from the rest precipitation products. The product also yields consistent performance with the small-scale watershed of Gilgel Abbay. This indicates that MSWEP under RainFact calibration mode is also

applicable for water resources application for medium scale of data scarce region in the upper Blue Nile basin, Ethiopia.

Table 6. Daily performance of various precipitation products of Kessie.

		Precipitation Products										
		Gauge Rainfall	CMORPH		TRMM		ERA-Interim		MSWEP		GPCP	
			S	R	S	R	S	R	S	R	S	R
NSCE	Cal	0.74	0.61	0.62	0.66	0.67	-2.6	-0.38	0.71	0.72	0.69	0.71
	Val	0.62	0.65	0.64	0.61	0.60	-1.7	-0.14	0.57	0.60	0.60	0.62
Bias	Cal	-7.5	-10.9	-16.7	-12.4	-16	97.1	44	-27.7	-18	-25.2	-18
	Val	-0.31	-18.8	-22	-25.6	-29	74.7	27	-30.3	-27	-32.9	-26
CC	Cal	0.87	0.81	0.81	0.83	0.83	0.86	0.86	0.87	0.87	0.86	0.86
	Val	0.80	0.82	0.82	0.81	0.81	0.76	0.77	0.79	0.80	0.81	0.81

Where, S= Simulation mode, R= RainFact calibration mode, Cal= Calibration period (2000-2007), Val= Validation period (2008-2012).

Both satellite and reanalysis (Except ERA-Interim) precipitation products have excellent performance (NSCE>0.75) in large scale (Eldiem) compared with small and medium scale of the upper Blue Nile basin in the calibration period for both simulation modes. This establishes that the distributed hydrological model CREST and the global precipitation product performances increase as the watershed area increases. As distributed hydrological model, the model error decreases with the increase of the drainage area [Xinyi Shen and Hong, 2014]. The simulated flow from TRMM performs excellent in terms of NSCE=0.86 evaluation as well, GPCP =0.79, CMORPH=0.78 and

MSWEP=0.76 also show very good results in the calibration period under simulation mode (S) as shown in Table 7. The same with Gilgel Abbay (no Lake Tana effect) and opposite to Kessie (to Lake Tana effect) the newly introduced RainFact calibration (R) mode shows high improvement on the reanalysis precipitation products (MSWEP and GPCC) in NSCE evaluation of Eldiem that has insignificant effect of Lake Tana. Especially, MSWEP shows the highest performance of NSCE=0.89 for both calibration and validation periods under the RainFact calibration (R) mode. The satellite precipitation products don't show any improvement in the application of RainFact calibration (R) mode over the simulation mode (S) in the Eldiem case study of NSCE evaluation.

Like Kessie (Medium scale) ERAI shows poor skill and needs improvement to capture the observed flow with NSCE value of -1.3 in calibration period under simulation mode (S). Resembling the other case studies of small and medium scale basin, the bias evaluation shows that minimum value of bias is observed in satellite simulated flow of CMORPH (-11.2%) over the reanalysis products under simulated mode (S) in large basin scale (Eldiem) as shown in Table 7. The RainFact calibration (R) mode of Eldiem shows that the simulated flow from TRMM shows the minimum value of bias (-9.9%) for the calibration period. This confirms that the bias statistics evaluation of satellite precipitation products shows high performance over the reanalysis precipitation products for both simulation modes for the large basin scale of Eldiem in the upper Blue Nile basin of the data scarce region.

Eldiem coefficient of correlation (CC) evaluation shows high performance for both calibration and validation periods under RainFact calibration (R) and simulation (S) modes. Like the case studies of Kessie and Gilgel Abbay MSWEP shows the best CC performance of 0.96 for both calibration and validation periods under RainFact calibration mode. These results of Eldiem show that introducing the newly RainFact calibration mode (R) shows high improvement for reanalysis

products. The same with the small and medium spatial scale case studies, MSWEP shows high performance of NSCE and CC that shows the product of MSWEP is a priori alternative data sources for water resources applications for the large basin scale of Eldiem, if precipitation adjustment is introduced.

Table 7. daily performance of various precipitation products of Eldiem.

		Precipitation Products										
		Gauge	CMORPH		TRMM		ERA-Interim		MSWEP		GPCP	
			S	R	S	R	S	R	S	R	S	R
NSCE	Cal	0.91	0.78	0.78	0.86	0.86	-1.3	0.57	0.76	0.89	0.79	0.86
	Val	0.93	0.75	0.75	0.74	0.76	-0.37	0.81	0.74	0.89	0.72	0.83
Bias	Cal	0.09	-11.2	-11.2	-13	-9.9	83.7	15.8	-35	-12	-31.5	-13.4
	Val	-6.2	-34	-34	-34.5	-32	50.5	-7	-37.3	-15.6	-36.5	-19.6
CC	Cal	0.96	0.89	0.89	0.93	0.93	0.92	0.92	0.95	0.96	0.93	0.93
	Val	0.97	0.93	0.93	0.95	0.95	0.96	0.90	0.95	0.96	0.93	0.93

Where, S= Simulation mode, R= RainFact calibration mode, Cal= Calibration period (2000-2007), Val= Validation period (2008-2012).

The result of this study shows that the simulated flow forcing with both satellite and reanalysis products perform well to capture the observed flow under simulation mode (S). However, the product from ERA-Interim shows poor skill and needs improvement to perform like the other products under simulation mode (S) especially for Kessie and Eldiem case studies. The reanalysis product of MSWEP shows consistent and very good performance based on NSCE and CC performance evaluation for all three case studies for both RainFact calibration (R) and simulation (S) modes in

the basin. The newly introduced RainFact calibration mode shows significant improvement for Gilgel Abbay and Eldiem case studies that have less reservoir effect, and Kessie (to Lake Tana effect) shows slight improvement for the RainFact calibration mode (R) for both satellite and reanalysis precipitation products.

The bias values reveal that the simulated flows have negative values, indicates that all products have limitation to capture the peaks and the total volume of observed flows. Satellite products show minimum bias value and better performance to capture the observed flow over the reanalysis products for both RainFact calibration (R) and simulation mode (S). The newly released reanalysis products don't show any improvement regarding on bias performance to capture the total volume of the runoff over the satellite products for different spatial scale case studies in the upper Blue Nile basin. The product from ERAI shows positive values of bias and less performance with high overestimation of the observed flow and shows need improvement to have better performance like the other products regarding on both NSCE and bias performance evaluations, applying for simulation mode (S).

The Quantile-Quantile (Q-Q) plot

Figure 18 shows the daily Q-Q plots of the simulated flow of various satellite and reanalysis precipitation products with a reference line of the observed flow for the three case studies in upper Blue Nile basin for the two RainFact calibration and simulation modes. The result of the Q-Q plot of the simulation mode for small scale basin study of Gilgel Abbay watershed shows that, following to the simulated flow from gauged rainfall, the simulated flow from CMORPH and MSWEP show better performance to capture the observed flow as shown in Figure 18a. The product from ERAI overestimates and shows poor performance relative to the other products under

the simulation mode. The RainFact calibration mode of Gilgel Abbay shows high improvement for the whole precipitation products and outperforms over the simulation mode as shown in Figure 18b. This indicates that applying RainFact calibration mode gives high improvement and efficiency over the simulation mode for the quantile flow estimation for water resources applications in the upper Blue Nile basin for small spatial scale case study of Gilgel Abbay watershed.

For medium spatial case study (Kessie), with the exception of ERAI, the whole global datasets of both satellite and reanalysis products show good performance to capture the reference line under the simulation mode as shown in Figure 18c. The RainFact calibration mode shows improvement for only reanalysis product of MSWEP and GPCC over the other products in the upper Blue Nile basin of medium scale of Kessie as shown in Figure 18d. This shows that applying RainFact calibration mode is only applicable and shows improvement for MSWEP and GPCC precipitation products over the simulation mode to estimate the quantile flow for the application of water resources applications in the upper Blue Nile basin for medium scale of Kessie.

The Q-Q plot of the simulation mode for the large spatial scale of Eldiem manifests that the satellite precipitation products of CMORPH and TRMM show better performance relative to the reanalysis products as shown in Figure 18e. Corresponding to the other case studies, the product from ERAI shows poor performance to estimate the quantile flow for large spatial scale of Eldiem under simulation mode. The RainFact calibration mode shows high improvement and it is more applicable for the reanalysis products of MSWEP and GPCC to estimate the quantiles over the simulation mode of the large basin scale of Eldiem in the upper Blue Nile basin as shown in Figure 18f.

The result of Q-Q plots of this study confirms that both reanalysis and satellite precipitation products perform differently for different case studies under simulation mode. For instance, MSWEP and CMORPH show better quantile estimation for Gilgel Abbay, for Kessie except ERAI the whole products perform well, as well for Eldiem satellite products of CMORPH and TRMM outperform over the other products under simulation mode. However, for RainFact calibration mode the reanalysis products of MSWEP and GPCC show consistent improvement and high performance to estimate the quantile observed flow of each case study over the other products. ERAI shows poor performance for all case studies for both simulation and RainFact simulation modes to estimate the quantiles of observed flow and it needs some improvement to perform like the other global precipitation products that have been considered in this study. These all results show that applying RainFact calibration mode for reanalysis products of MSWEP and GPCC gives high improvement and performance to estimate the observed quantile flow over the simulation mode at daily temporal scale. Especially, MSWEP can be an alternative source of data to estimate the quantile flow for water resources planning and management in data scarce areas of Africa without applying specific calibration for each precipitation product.

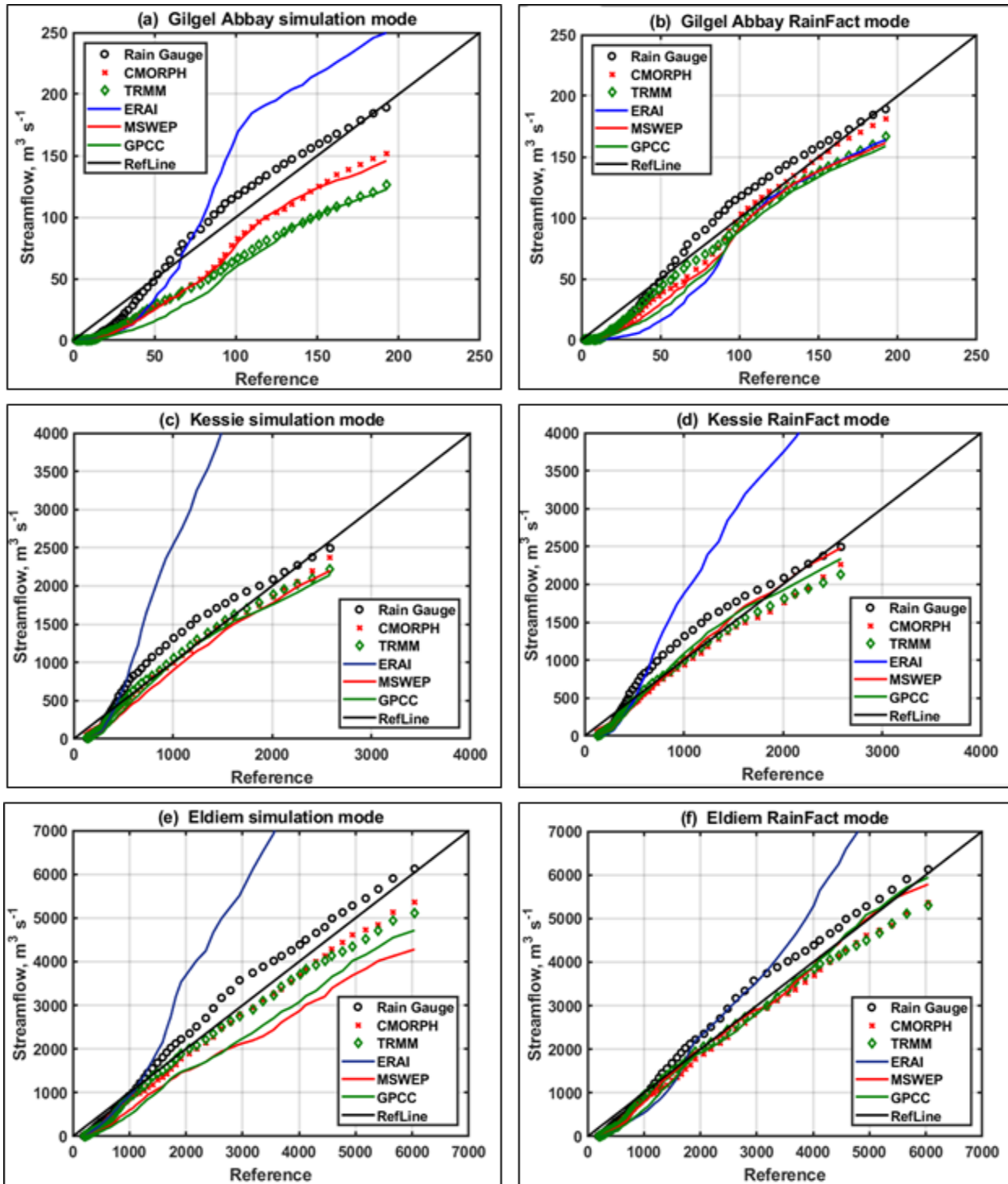


Figure 18. Daily Quantile-Quantile plots of Gilgel Abbay (a) simulation mode (b) RainFact calibration mode, Kessie (c) simulation mode (d) RainFact calibration mode, Eldiem (e) simulation mode (f) RainFact calibration mode

4.3.1. Error of precipitation propagation to the simulated flow

Proportionality of error of remotely sensed (global satellite and reanalysis) precipitation products to the magnitude of gauged rainfall, and the simulated flow of global precipitation products to the magnitude of observed flow is investigated in this study using bias (equation 4). The magnitude and the bias sign used to characterize the error propagation of the global precipitation products to the simulated flow for the three nested watersheds.

The result of the global precipitation products to the gauged rainfall bias, and the simulated flow of the global precipitation products to the observed flow bias is shown in Table 8. The result indicates that the magnitude and the bias sign of the precipitation products to the gauged rainfall, and the simulated flow of the global precipitation products to the observed flow are consistent in a certain direction and in replicating the magnitude proportions for the three case studies. This indicates that the error of both satellite and reanalysis global precipitation products to the simulated flow can be characterized as systematic error for the three nested watersheds in the upper Blue Nile basin. Since the systematic error is a series of errors in accuracy that are consistent in a certain direction, while random errors are those which are caused by random and unpredictable variation in an experiment [Amir *et al.*, 2012]. Generally, systematic error is introduced by a problem that is consistent through an entire experiment.

For the small scale of the basin of Gilgel Abbay minimum bias value of precipitation is observed in CMORPH with an underestimation of the gauged rainfall with -8.8% bias value and the only precipitation product of the ERAI overestimates the gauged rainfall with 28.3% bias. The same result is replicated for the simulated flow, minimum bias is observed in the CMORPH with an underestimation of the observed flow with -21.9% bias and the simulated flow from ERAI overestimates the observed flow with 24.4% bias value as shown in Table 8. This indicates that

the error magnitude and sign of the precipitation products to estimate the gauged rainfall propagates to the simulated flow, and the error can be characterized as systematic error for the small scale of the Gilgel Abbay watershed.

Minimum bias value with an underestimation of the gauged rainfall is observed in the CMORPH precipitation product with -2.4% bias value in the medium scale of Kessie, and ERAI overestimates with high bias value of 56.6% to estimate the gauged rainfall. The same with the Gilgel Abbay watershed, the error of the precipitation to estimate the gauged rainfall propagates to the simulated flow in the medium scale of basin case study of Kessie, and the error is characterized as systematic error with consistent magnitude proportion and sign of bias value.

The same with Gilgel Abbay and Kessie, the error of the precipitation products to estimate the gauged rainfall of the large basin scale of Eldiem propagates to the simulated flow with consistent bias magnitude proportions and sign. These all bias results of the three different sized case studies show that the performance of the global precipitation products to estimate the gauged rainfall has high impact on the performance of the simulated flow in the upper Blue Nile basin. The error of the global precipitation products propagates to the simulated flow and the error can be characterized as systematic error as the bias magnitude proportions and sign are consistent to estimate the gauged rainfall and to simulate the observed flow.

Table 8. Error (bias %) of satellite and reanalysis precipitation products propagates to the simulated flow from 2000 to 2012.

Gilgel Abbay						
		CMORPH	TRMM	ERA1	MSWEP	RGPC
Precipitation	Bias (%)	-8.8	-13.2	28.3	-12.4	-20.9

Runoff	Bias (%)	-21.9	-42.0	24.4	-29.4	-40.5
Kessie						
		CMORPH	TRMM	ERA-Interim	MSWEP	GPCC
Precipitation	Bias (%)	-2.4	-3.7	56.6	-4.2	-5.3
Runoff	Bias (%)	-10.7	-12.8	95.6	-28.0	-25.5
Eldiem						
		CMORPH	TRMM	ERA-Interim	MSWEP	GPCC
Precipitation	Bias (%)	-2.5	-2.6	41.9	-6.4	-3.4
Runoff	Bias (%)	-11.4	-13.3	83.1	-35.1	-31.7

4.4. Conclusive Remarks

The main purpose of this study is comparison and evaluation of globally available satellite and reanalysis precipitation products over a multi-scale (1,656–199,812 km²) and multi-year (2000–2012) for data scarce basin in Africa, upper Blue Nile basin (Abbay). The fully distributed hydrological model CREST is calibrated using in-situ observed rainfall data and used to evaluate the global precipitation products in simulation replacing the rain gauge forcing with each global precipitation product with the observed flow using the rain gauge calibrated model parameters. The watersheds included in the study are Gilgel Abbay (1,656 km²), Abbay at Kessie (65,784 km²) and Abbay at Eldiem (199,812 km²). Two satellite globally available precipitation products of gauge adjusted CMORPH, and TRMM TMPA 3B42v7, as well three reanalysis precipitation products of ERA-Interim, MSWEP and MSWEP with 0.25° spatial resolution are evaluated for daily simulated flow. The comparison has made between the simulation mode (S) using rain gauge calibrated parameters, and RainFact calibration mode (R) applying rain factor adjustment on the

RainFact parameter for each product to improve the performance of the product for water resources application.

The performance of the model applying gauge-based rainfall data are evaluated using three coefficient of Nash-Sutcliffe Efficiency (NSCE), Bias and Coefficient of Correlation (CC). The calibration result of NSCE for each basin scale shows that, the fully distributed hydrological model CREST is applicable for multi-scale of data scarce region of Abbay basin (upper Blue Nile basin) ranges from 0.74 to 0.91 for calibration period and from 0.62 to 0.93 for validation period.

Based on NSCE and CC evaluation criteria of the simulated flow forcing the global precipitation datasets applying the rain gauge calibrated model parameters (simulation mode) results indicate that the hydrological performance of both satellite and reanalysis products (except ERAI) provide encouraging results for all spatial scales of NSCE (0.57 to 0.86) in calibration period. The reanalysis product of MSWEP shows consistent and very good performance for all three case studies of both RainFact calibration (R) and simulation (S) modes in the basin. The newly introduced RainFact calibration mode shows significant improvement for Gilgel Abbay and Eldiem case studies that have less reservoir effect and for Kessie (to Lake Tana effect) shows modest improvement for the RainFact calibration mode (R) for both satellite and reanalysis precipitation products on NSCE and CC evaluation criterion.

The bias value evaluation reveals that the simulated flows have negative values, indicates that all products have limitation to capture the peaks and underestimate the total volume of observed flows. Satellite products show minimum bias value to capture the observed flow over the reanalysis products for both RainFact calibration (R) and simulation modes (S). The newly released reanalysis products don't show any improvement regarding on bias performance to capture the total volume

of the runoff over the satellite products for different spatial scale case studies in the upper Blue Nile basin.

The result of QQ plots of this study confirms that both reanalysis and satellite precipitation products perform differently for different case studies under simulation mode. However, applying RainFact calibration mode for reanalysis products of MSWEP and GPCC shows high improvement and gives consistent performance to estimate the observed quantile flow over the simulation mode at daily temporal scale.

The total performance evaluation of NSCE, CC, Bias and Q-Q plot show that the product from MSWEP with the RainFact calibration mode shows high performance and shows high improvement over the simulation mode from the rest precipitation products. The MSWEP reanalysis precipitation product can be a priori alternative source of data for various water resources applications in the upper Blue Nile basin for its predictive ability and the length of data availability which reanalysis products have advantage over the satellite products.

This study considers only three different sized watersheds for the analysis. Further investigation is desirable to contemplate more watersheds on the basin with high quality of datasets.

Chapter Five

5. Evaluation of Global Water Resources Reanalysis Runoff Products for Local Water Resources Applications: A case of Upper Blue Nile Basin in Ethiopia.

5.1. Introduction

The combined effects of growing population, rising incomes, and expanding cities will non-linearly increase the water demands, while supply becomes more erratic and uncertain due to climate variability. World Bank estimates that with current population growth and water management practices, the world will face a 40% shortfall between forecast demand and available water supply by 2030. Feeding nine billion people by 2050 will require a 60% increase in agricultural production and a 15% increase in water withdrawals. More than half of the world's population now lives in urban areas and this number is growing fast, increasing water withdrawal. By 2025, it is estimated that about 1.8 billion people will be living in regions or countries with absolute water scarcity. Water scarcity is further exacerbated by climate change and the combined effects could cost some regions up to 6% of their GDP, spur migration, and spark conflict (<http://www.worldbank.org/en/topic/water/overview>). Water resources planning, development and management for sustainable socio-economic development requires understanding the spatial and temporal availability of the resources at multiple scales (regional, national and sub-national level). This is particularly important in developing regions where population and urbanization are growing rapidly, while limited availability of hydrological data hinders sound water development and management. This research focuses on improving the quality and availability of local data by using globally accessible hydrological data products.

The availability and quality of climate data varies from region to region, and observations are sparse and irregularly distributed in many areas [*El-Sadek et al.*, 2011; *Haberlandt and Kite*, 1998; *Kanamaru and Kanamitsu*, 2007]. In Africa, gauging stations are typically located nearby towns or vehicle accessible sites. The vast rural basins remain either ungauged and the quality of the data is compromised. Lakew et al., 2017 compared the spatial coverage of specific hydrological data of African climate stations to stations in Germany. There are about 1150 streamflow stations in Germany, while in sub-Saharan Africa, covering 60 times larger area, there are only 888 streamflow gauging stations. The in situ data situation is similar in Ethiopia. To overcome the decline of data availability in developing regions, and the keen interest to generate a consistent set of hydrological data products for global climatic and environmental understanding, alternative sources of data have been emerging from different research institutions globally. A team of researchers in Europe in the framework of earthH2Observe project have recently generated a global water resources reanalysis based on combination of multiple hydrologic models a global atmospheric reanalysis and earth observation datasets [*H Beck et al.*, 2017; *Schellekens et al.*, 2017].

Atmospheric reanalysis provides a reconstruction of the historical global atmospheric circulation patterns through assimilation of available observational data into Global Climate Model (GCM) simulations. Reanalyses data are widely considered to be a robust proxy of historical atmospheric observations that reproduce observed climate dynamics and are therefore useful for climate and hydrological application studies where there are inadequate meteorological input data [*Hwang et al.*, 2013]. Water resource reanalysis (WRR) in earthH2Observe consisted of forcing multiple hydrological models with atmospheric reanalysis data. One of the WRR generated data is runoff, which is important for areas where in-situ streamflow data are non-existent or limited. Although reanalysis products hold a great potential for studies and applications related to water resources

their level of accuracy needs to be evaluated at local scale. The WRR products consist of errors cascaded from precipitation forcing, model formulation and processing simplification. Recently [H E Beck *et al.*, 2017] run 10 global hydrological models for over 900 medium size catchments and revealed the importance of both model and input data uncertainty, which needs to be carefully evaluated before applying global hydrological model driven by global scale precipitation data at local scale.

In this study, we focus on the performance evaluation of two earth2Observe Water Resources Reanalysis river runoff products, WRR1 available at 0.5° and WRR2 available at 0.25° used WFDEI and MSWEP forcing data respectively, in terms of flow variability characterization in the upper Blue Nile basin. These WRR runoff products are compared to in situ observed runoff data and contrasted to runoff simulated using a locally calibrated distributed hydrological model (Coupled Routing and Excess Storage; CREST v2.1) [Xinyi Shen and Hong, 2014] forced with in situ local as well as the global precipitation dataset (MSWEP) used in WRR2. The runoff simulations are evaluated for three basin scales in the upper Blue Nile: Gilgel Abbay, representing a small-scale basin ($1,656\text{km}^2$); Kessie, representing a medium size basin ($65,784\text{ km}^2$) and the entire upper Blue Nile at Eldiem ($199,812\text{ km}^2$), based on multi-year observed streamflow data (2000-2012). Evaluation is performed at daily, monthly and for the wet season (June to September, JJAS) temporal scales. Different statistical error metrics and quantile-quantile (Q-Q plot) scatter plots are used to evaluate the quality of these products for representing the upper Blue Nile basin flow variability at different scales. This evaluation inform uses of the WRR data for management of the basin's water resources (hydropower, irrigation and water supply).

5.2. Data and Methods

5.2.1. Study Area

This study is conducted in the upper Blue Nile basin in three different sized watersheds, Gilgel Abbay (small), Abbay at Kessie station (middle upper Blue Nile) and the upper Blue Nile at Eldiem (at the outlet of the basin and border to Sudan) (Figure 19).

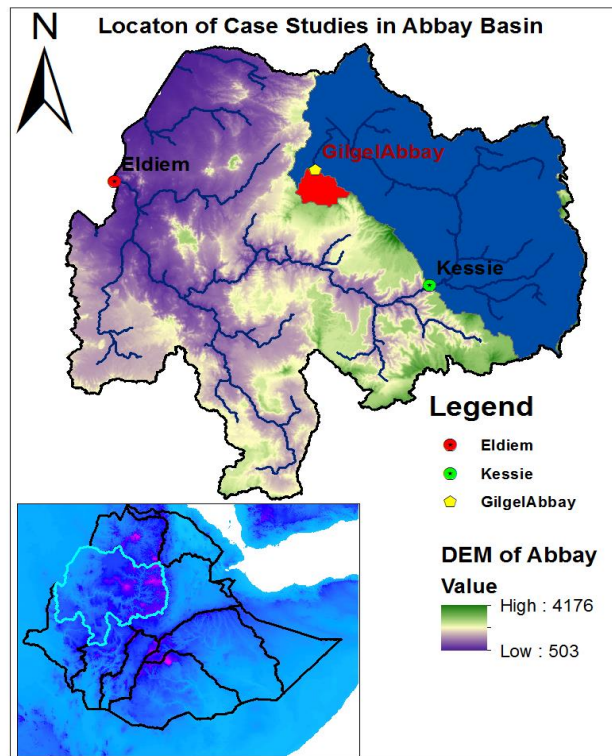


Figure 19. Location of the three case studies in upper Blue Nile basin.

5.2.2. Data Sets

The daily and monthly temporal scales of observed streamflow multi-year 2000-2012 dataset of the three case studies are collected from the Ethiopian Ministry of Water, Irrigation and Electricity. From the whole upper Blue Nile basin, the three stations are selected attributable to the quality and

the availability of updated data as well stations that can show the three spatial scales of small, medium and large in the basin.

Water resources reanalysis river runoff (WRR1, and WRR2) products with daily and monthly temporal resolutions are taken from the earth2Observe data portal (<https://wci.earth2observe.eu/portal/>). The global river runoff products that are considered in this study are categorized and released by the first and second rounds. The first round (WRR1) of earth2Observe released seven products with spatial resolution of 0.5° and used WFDEI forcing data (CNRS, CSIRO, ECMWF, JRC, Meteo France, Universitat Kassel and Universiteit Utrecht). The second round (WRR2) of the project released four reanalysis products (ECMWF, JRC, Meteo France, and Universitat Kassel) that have a spatial resolution of 0.25° and used MSWEP forcing data.

For this study the daily temporal scale precipitation (MSWEP) is used for the simulation of daily runoff using the locally set up distributed CREST hydrological model to compare with global model runoff products. The CREST is calibrated using the gauged rainfall and the gauged based calibrated parameters are used to simulate the global precipitation MSWEP forcing.

The NSCE performance evaluation is carried out for the daily and monthly runoff computations that have a relatively good length of data. For seasonal computation, we omitted the NSCE efficiency evaluation due to the length of years is limited from 2000 to 2012. The seasonal runoff prediction is evaluated using the Correlation Coefficient (CC), Root Mean Square Difference (RMSD) and Standard Deviation (S).

For evaluation of monthly and seasonal temporal scales of WRR2 products as well the simulated runoff from CREST model forced with MSWEP, we aggregate the daily time scale into monthly and seasonal runoffs.

5.3. Result and Discussion

Two Earth2Observe Water Resources Reanalysis river runoff products, WRR1 available at 0.5° and WRR2 available at 0.25° used WFDEI and MSWEP forcing data respectively are compared to in situ observed runoff data and runoff simulated using a locally calibrated CREST distributed hydrological model forced with the global precipitation MSWEP dataset used in WRR. The results are compared at three spatial and temporal scales using different statistical and graphical methods and presented below.

The NSCE performance and the Taylors Diagram

The first version of WRR1, which has coarser spatial resolution of reanalysis river runoff products, shows differing and inconsistent performance at different temporal and spatial scales of the three nested watersheds. No single product performs consistently better than the other products at both daily and monthly temporal scales. For the small-scale basin of Gilgel Abbay, monthly evaluation of NSCE shows that JRC (0.79), Universitat Kassel (0.73) and CNRS (0.67) products perform better compared with the other WRR1 products. In terms of daily temporal scale, JRC performs better than all other products with NSCE of 0.65. CSIRO shows the poorest skill and clearly overestimates the observed runoff with little improvement even at monthly time scale. The NSCE value at monthly temporal scale remains negative. Among the different WRR1 products, JRC shows better and more consistent performance based on NSCE statistics for monthly (0.79) and daily (0.65) temporal scales over Gilgel Abbay watershed which represents small spatial scale as shown in Table 9. Compared to the runoff simulated by using locally set CREST hydrological

model and forced with in situ gauge observed rainfall as well as global precipitation product, the latter estimates have a better edge than the best WRR1 product which is the JRC.

Similarly, the high resolution WRR2 runoff products give differing results. Unlike the coarser WRR1 runoff products, we found the Universitat Kassel (UniK) product improves from its predecessor WRR1 product significantly, while JRC product shows decrease in its performance. The NSCE value of the (WRR2) UniK product improved from 0.19 (WRR1) to 0.62 (WRR2) and from 0.73 (WRR1) to 0.76 (WRR2) for the daily and monthly temporal scales respectively, while the JRC decreased from 0.65 (WRR1) to 0.55 (WRR2) and from 0.79 (WRR1) to 0.66 (WRR2) for the daily and monthly temporal scales respectively.

The results of locally set distributed CREST model simulation runoff forced with the global precipitation product (MSWEP) has shown a better performance than all global WRR runoff products at Gilgel Abbay (Table 9).

Table 9. NSCE performance of WRR products and simulated runoff of Gilgel Abbay watershed (small-scale 1,656 km²).

WRR products	WRR1 (0.5°*0.5°)		WRR2 (0.25°*0.25°)	
	From 2000 to 2012		From 2000 to 2012	
	Daily	Monthly	Daily	Monthly
CREST/Gauge			0.76	0.88
CREST/MSWEP			0.71	0.81
ECMWF	-0.47	0.49	0.32	0.46

JRC	0.65	0.79	0.54	0.66
Meteo France (MeteoFr)	0.37	0.53	0.25	0.34
Universitat Kassel (UniK)	0.19	0.73	0.62	0.76
CNRS	0.06	0.67		
CSIRO	-4.47	-3.82		
Universiteit Utrecht (UniUt)	0.03	0.11		

The performance of the products evaluated using Root Mean Square Difference (RMSD), Coefficient of Correlation (CC) and Standard Deviation (S) is presented using Taylor diagram [Taylor, 2001]. Taylor diagrams are mathematical diagrams designed to graphically indicate which of several approximate representations (or models) of a system, process, or phenomenon is most realistic. This diagram facilitates the comparative assessment of different models. It is used to quantify the degree of correspondence between the modeled and observed behavior in terms of three statistics. The nearest point that represents the global model runoff product to the reference (Ref) shows the higher and better performance to estimate the observed flow. The different colors indicate different products, the red and black colors represent the coarser WRR2 and the finer WRR2 global runoff products respectively. The blues are the simulated flow from the global precipitation product and from the gauged rainfall using CREST model. Overall, the diagram shows that the high-resolution products from WRR2 (ECMWF, JRC and Universitat Kassel) exhibit improvement relative to the coarser WRR1 products, particularly for the seasonal runoff of the small basin (see Figure 20c). From the WRR2 products, the global products obtained from

ECMWF and UniK have shown comparative performance at daily, monthly and seasonal scales. While the performance of UniK product is consistent with the NSCE performance matrix, the ECMWF product appears to display comparative sometime better visual performance (at seasonal scale) than the UniK product. This may be attributed to systematic bias and larger relative difference between the unexplained and the initial variance which undermines the NSCE of ECMWF performance but provides appealing visual in the Taylor diagram for ECMWF product. Though the (WRR2) UniK and ECMWF runoff product stand out from its counterpart models at Gilgel Abbay, the simulated runoff from the locally set CREST hydrological model forced with the global precipitation product of MSWEP has relatively a better edge than the earlier products at daily and monthly scales.

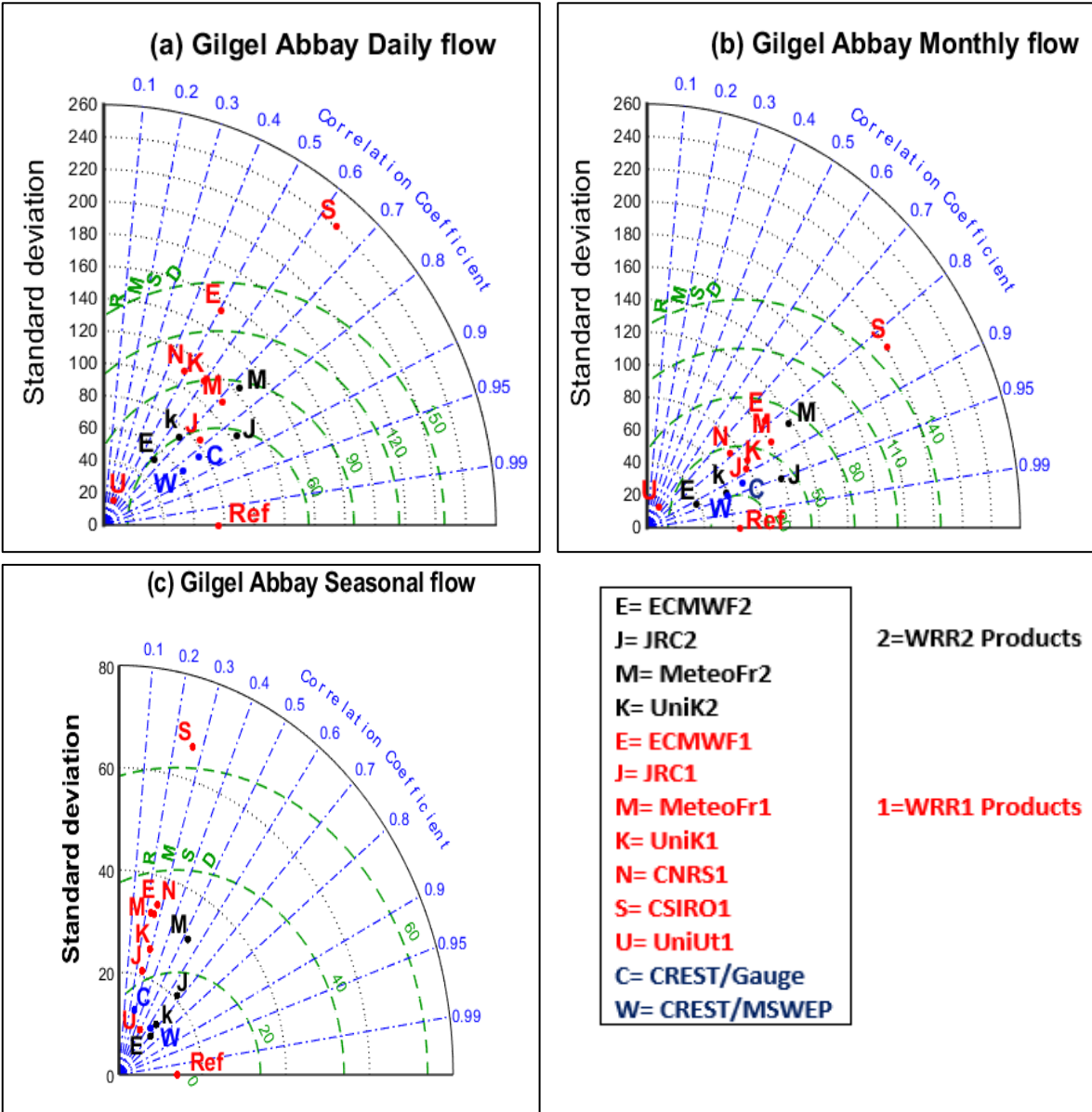


Figure 20. Taylor diagram of (a) Daily, (b) Monthly and (c) Seasonal runoff of Gilgel Abbay watershed.

The evaluation of WRR1 reanalysis river runoff products based on NSCE metric for Kessie station (medium scale) at monthly temporal scale shows that CNRS (0.77), Universitat Kassel (0.66), CSIRO (0.57) and Meteo France (0.56) products perform better. The products ECMWF (0.35) and JRC (0.34) performs relatively poorly, and the Universiteit Utrecht product shows the poorest skill (NSCE = -0.31). As in the Gilgel Abbay case study, the daily temporal scale in the Kessie case

study shows relatively poorer performance than at the monthly scale. The CNRS product gives better performance than all other products (NRCE = 0.64) at daily scale. In contrast to the small basin scale, the CNRS runoff product gives a better skill for both daily and monthly temporal scales than the rest of the WRR1 products (see Table 10).

Like Gilgel Abbay watershed, the WRR2 high resolution products the UniK product shows higher improvement than the WRR1 products and relative to the other WRR2 products. The NSCE performance of UniK product at daily and monthly scales is 0.66 and 0.83 respectively followed by the MeteoFr with daily and monthly NSCE performance of 0.54 and 0.68 respectively.

In contrast to Gilgel Abbay, the daily and monthly NSCE performances of runoff estimated with locally set hydrological model (CREST) show slightly lower value as compared to the WRR2 global UniK runoff product. We noted that this may be imputable to the effect of Lake Tana that has a reservoir surface area of ~3,500 km² located upstream of Kessie observation station. The distributed hydrological model (CREST) has a limitation of considering the storage effects of upstream lakes and swamps on hydrological performance. The performance of Universitat Kassel (WRR2) and CNRS (WRR1) products show similar NSCE skill with the locally set CREST model at both daily and monthly temporal scales.

Table 10. NSCE performance of WRR products and simulated flows of Kessie (medium-scale 65,784 km²).

WRR products	WRR1 (0.5°*0.5°)		WRR2 (0.25°*0.25°)	
	From 2000 to 2012		From 2000 to 2012	
	Daily	Monthly	Daily	Monthly

CREST/Gauge			0.66	0.80
CREST/MSWEP			0.65	0.77
ECMWF	0.03	0.35	0.27	0.70
JRC	0.31	0.34	0.26	0.33
Meteo France (MeteoFr)	0.49	0.56	0.54	0.68
Universitat Kassel (UniK)	0.34	0.66	0.66	0.83
CNRS	0.64	0.77		
CSIRO	0.43	0.57		
Universiteit Utrecht (UniUt)	-0.26	-0.31		

The Taylor diagram of the medium scale Kessie shows that based on RMSE, CC and S statistical performance evaluation, for the daily and monthly runoff computations, the simulated runoff from the locally set hydrological model CREST do not exhibit any significant difference to the performance of the global WRR1 products from hydrological models of CNRS and Meteo France. The recently released WRR2 products also show weaker performance relative to the coarser WRR1 products at Kessie. The seasonal runoff products at Kessie show weaker performance that may imputable to the Lake Tana reservoir effect as described above and shown in Figure 21c.

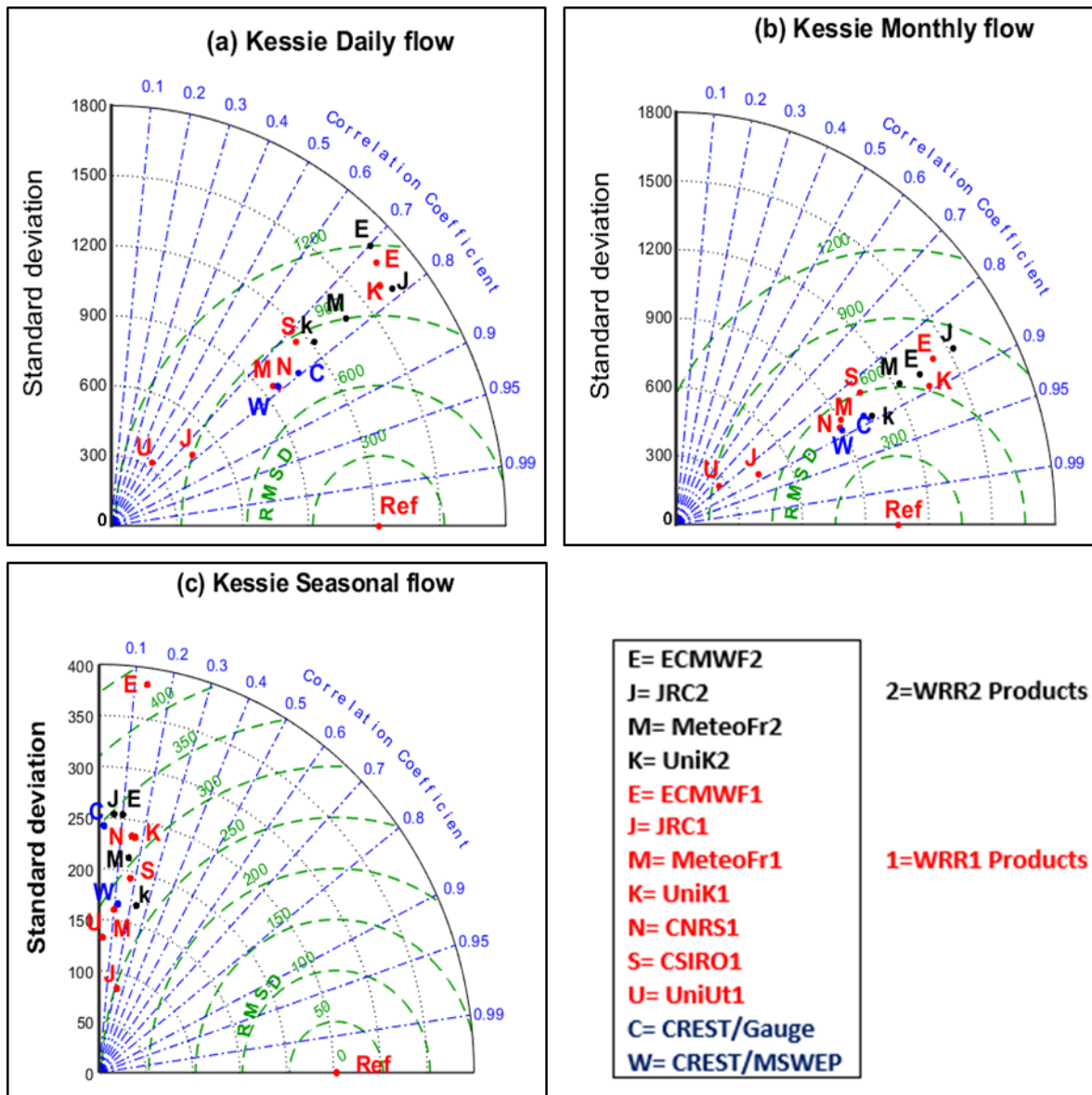


Figure 21. Taylor diagram of (a) Daily, (b) Monthly and (c) Seasonal runoff of Kessie.

As with the small and medium size basins, the WRR1 products perform inconsistently and relatively poorly in capturing the observed runoff at the outlet of the Blue Nile at Eldiem station, for both monthly and daily temporal scales except for the CNRS product performing better at both daily and monthly temporal scales. The CNRS performs well with NSCE of 0.75 and 0.86 at both daily and monthly time scales respectively. While UniK product performance with NSCE of 0.58

at monthly time scale, all the rest of the products perform poorly for all WRR1 products including the UniK daily product (Table 11).

The performance of WRR2 products for the entire upper Blue Nile basin is significantly improved for UniK product relative to the corresponding WRR1 products and from the rest of WRR2 products (Table 11). The monthly and daily NSCE matrix improve from 0.58 to 0.83 and from 0.36 to 0.82 respectively. The rest of WRR2 products performance is as poor as the WRR1 products. The performance of the simulated runoff from the locally set CREST model is significantly better than that of the reanalysis products at both monthly (NSCE = 0.95) and daily (0.91) scales. As Lake Tana is far from the outlet and its contribution is less significant at the upper Blue Nile scale, its effect is diminished. As a result, the locally set hydrological model CREST performance is high, in contrast to the Kessie basin that is affected by lake storage. The higher resolution WRR2 product of UniK exhibits comparative performance with the locally set hydrologic model CREST for both daily and monthly evaluations of NSCE.

Table 11. NSCE performance of WRR products and simulated flows of Eldiem (large-scale 199,812 km²).

WRR products	WRR1 (0.5°*0.5°)		WRR2 (0.25°*0.25°)	
	From 2000 to 2012		From 2000 to 2012	
	Daily	Monthly	Daily	Monthly
CREST/Gauge			0.91	0.95
CREST/MSWEP			0.76	0.78
ECMWF	-0.21	0.04	-0.22	-0.12

JRC	-2.30	-2.44	-2.60	-2.9
Meteo France (MeteoFr)	-0.37	-0.38	-2.1	-2.3
Universitat Kassel (UniK)	0.36	0.58	0.82	0.83
CNRS	0.75	0.86		
CSIRO	-0.53	-0.49		
Universiteit Utrecht (UniUt)	-0.26	-0.43		

As shown in the Taylor diagram (Figure 22), the CREST model results have better performance than the WRR products. However, the performance of the WRR2 products have higher performance than the coarser WRR1 products as observed in the other basins. The WRR2 products have a clear deviation from the WRR1 products in terms of the Correlation Coefficient (CC) as shown in Figure 22 for the daily, monthly and seasonal scales. The seasonal Taylor diagram (Figure 22c) and NSCE show that the WRR2 global product from UniK model is the best suited and potentially applicable for characterizing the upper Blue Nile hydrologic variability.

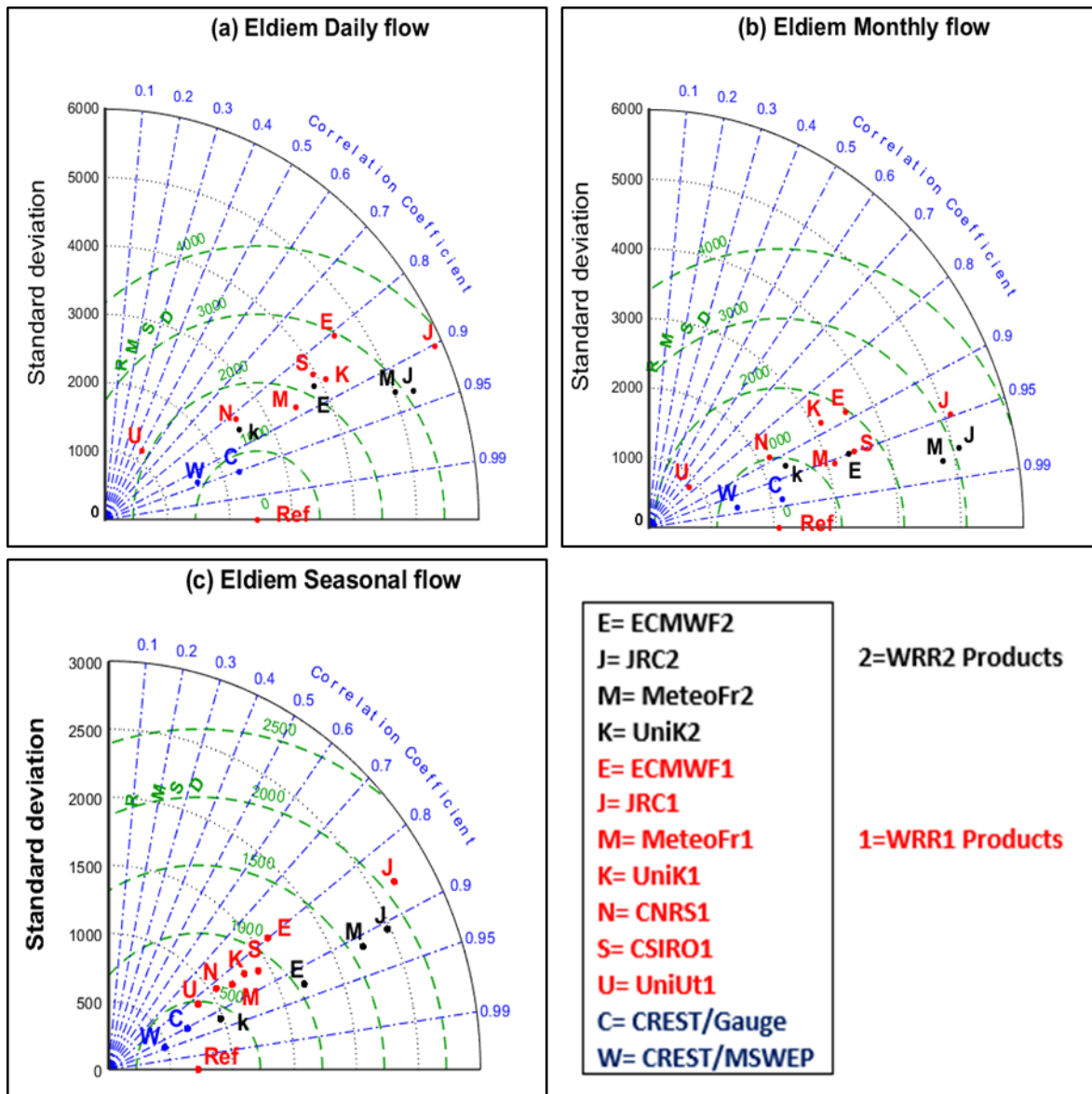


Figure 22. Taylor diagram of (a) Daily, (b) Monthly and (c) Seasonal runoff of Eldiem.

The comparison of the runoff simulations from the global hydrological models with the runoff simulations from the locally-set hydrological model (CREST) shows that the latter appears to have a better edge than the earlier at daily and monthly temporal scales. This is imputable to the locally set CREST model uses gauged rainfall and applies calibration by tuning the parameters to estimate the observed flow at a daily base. This makes the local model CREST outperform the global

products at daily and monthly spatial scales. At seasonal temporal scale, the best product of WRR2 (the UniK model) shows better performance than the locally set hydrologic model indicating that temporal scale has pronounced effect in evaluating hydrological model products performance for the three nested spatial scales of this study.

Based on the NSCE and Taylors diagram evaluations, all products of the WRR1 and WRR2 provide differing results at different spatial and temporal scales. The JRC (WRR1) product is better than the rest of WRR1 and WRR2 products at smaller scale followed with (WRR2) UniK product with comparative performance. At Kessie (medium scale), the UniK product provides a better performance than the rest of the products followed with (WRR1) CNRS product. At Eldiem (border, large scale), the (WRR2) UniK product provides superior performance followed with (WRR1) CNRS product and the monthly product of ECMWF product.

The UniK product shows better performance from the rest WRR products that is imputable to the model computes time-series of fast-surface and subsurface runoff, groundwater recharge and river discharge as well as storage variations of water in canopy, snow, soil, groundwater, lakes, wetlands and rivers.

Water resources planning and management requires different temporal scales depending on the purpose of the project. Seasonal products may be required for seasonal inflow forecasting to storage reservoirs. Whereas monthly products may be beneficial for reservoir release operation, the daily products may be required for flow forecasting for relatively smaller catchments. Depending on the requirement of temporal scale (daily, monthly, seasonal), different products may be used for range of water resources applications and purposes equally. However, the (WRR2) UniK global runoff product and the runoff product estimated by locally set CREST hydrological model forced with global MSWEP precipitation product consistently performs at all spatial and

temporal scales indicating the immediate usefulness of the two products for various water resources management products in the upper Blue Nile basin.

The Quantile-Quantile (Q-Q) plot

Further analysis of the Quantile-Quantile (Q-Q) is undertaken to understand the two datasets come from the same population group. Q-Q plot is a graphical technique for determining if two datasets come from populations with a common distribution [Wilk and Gnanadesikan, 1968]. The Q-Q plots of daily, monthly and seasonal scales for WRR1 products and for the three basin scales are plotted in Figure 23. Results reveal different distributions and performances for the different WRR1 products and scales.

For the small-scale of Gilgel Abbay, the WRR1 products perform well and are able to fall along the 45-deg line except for CSIRO and Universiteit Utrecht (UniUt) products. The CREST simulated runoff based on MSWEP performs worse than the WRR1 products of JRC and CNRS, for the three temporal scales as shown in Figure 23a, b& c. At the medium basin size, the Q-Q plots, Figures 23d, e&f, indicate that CNRS, CSIRO, and Meteo France products of WRR1 perform much better than the rest of WRR1 products as well as the CREST forced with MSWEP simulated runoff. At the upper Blue Nile basin, the Q-Q plots, Figures 23g, h&i, show that except for CNRS the rest WRR1 products including the CREST simulated runoff forced with MSWEP performed poorly; especially JRC exhibits strong overestimation and poor agreement.

Overall, the result of the Q-Q plots of WRR1 reveals inconsistent performance of the models for the three basin scales. However, for the medium and large size basins CNRS performs better than the rest WRR1 products. The CREST simulated runoff forced with MSWEP shows weaker performance than the CNRS (WRR1) product at all temporal scales of the case studies in the upper Blue Nile basin.

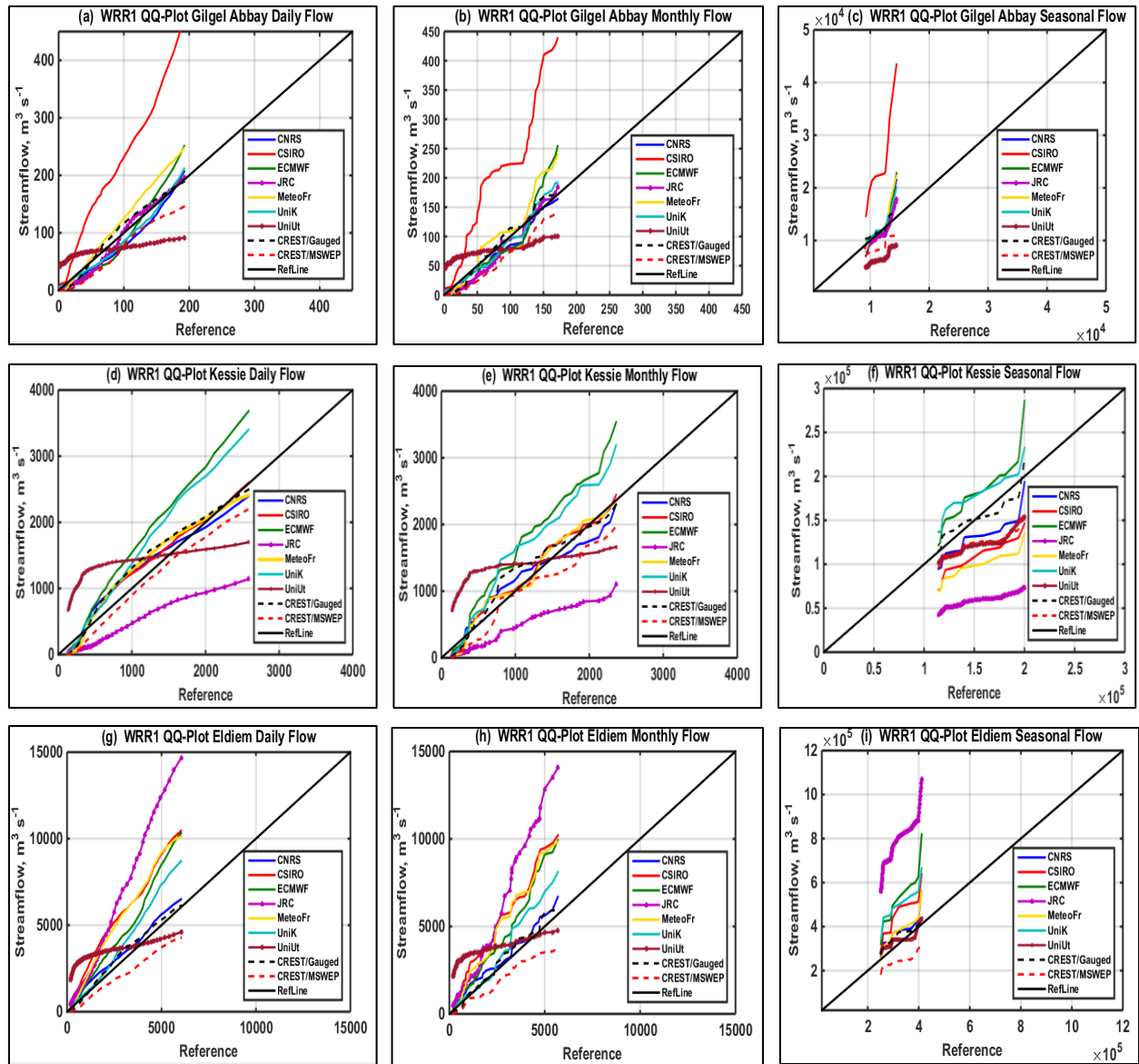


Figure 23. WRR1 QQ-plot of Gilgel Abbay watershed (a) Daily (b) Monthly (c) Seasonal, Kessie station (d) Daily (e) Monthly (f) Seasonal, Eldiem Station (g) Daily (h) Monthly (i) Seasonal runoff.

The Q-Q plots for the WRR2 products (Figure 24) show that the product of UniK captures very well the reference line for all temporal and spatial scales of the basin. Generally, the Q-Q plot shows significantly improved performance over the earlier WRR1 product. The simulated runoff

from CREST forced with MSWEP shows relatively poor performance relative to the Universitat Kassel product as shown in Figure 24.

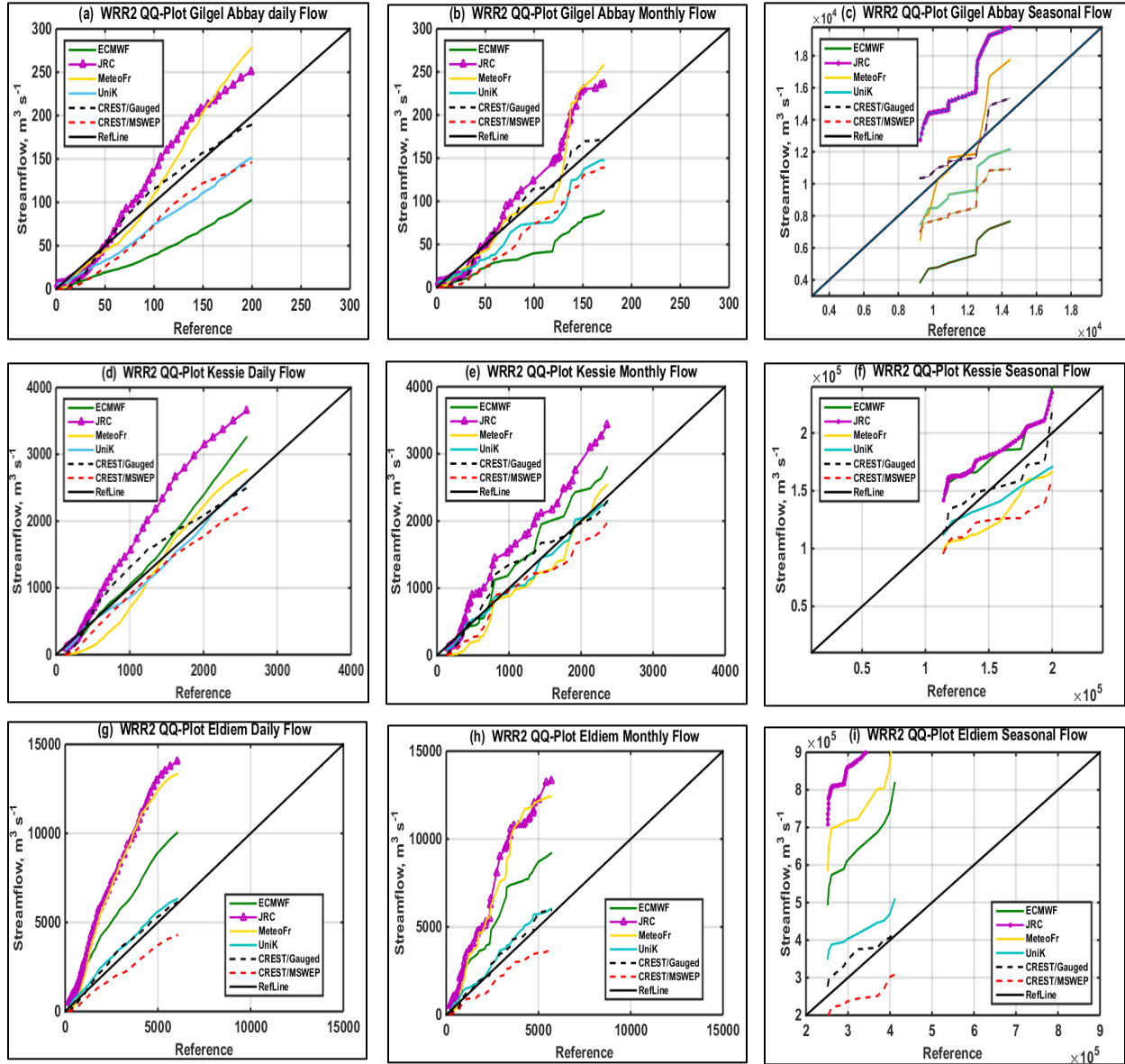


Figure 24. WRR2 QQ-plot of Gilgel Abbay watershed (a) Daily (b) Monthly (c) Seasonal, Kessie station (d) Daily (e) Monthly (f) Seasonal, Eldiem Station (g) Daily (h) Monthly (i) Seasonal runoff.

Critical Success Index (CSI)

As a final step of this evaluation, the WRR products and the simulated runoff from CREST are evaluated in terms of their ability to predict flows exceeding threshold values. Water management decisions (e.g. reservoir water releases or water allocations) are based on threshold values of water flows. In this study we use the Critical Success Index (CSI) as the metric to evaluate the models' performance in detecting flows exceeding the 50th and 80th-percentile of river flows for the three basin scales. CSI is determined from the two-dimensional contingency table defining the events where observed and forecasted flows exceed defined threshold [Donaldson. *et al.*, 1975; Doswell *et al.*, 1990].

In this study, the thresholds were the 50th-percentile of seasonal flow and the 80th-percentile of monthly flows. The 50th-percentile is used for the seasonal flows to evaluate the ability of the models to capture conditions above or below the long-term average. The 80th-percentile is used to evaluate the models' ability to detect low frequency flows. A hit (A) in the contingency table is defined as a yes/yes event, meaning a runoff greater than the percentile threshold is reported in both the observed and predicted runoff. A miss (C) is defined as a yes/no event, meaning a runoff greater than the percentile threshold is reported in observed flow, but it is not reported in the model simulations. A no/yes event is known as a false alarm (B), indicating that the model reports runoff exceeding the percentile threshold, but this is not confirmed by the observed flows. The information for no/no events, known as a correct null (D), is not collected. The Critical Success Index (CSI) is then defined as:

$$CSI = \frac{A}{(A + B + C)} \quad (7)$$

Based on the CSI performance evaluation, the best warning performance statistic of the model occurs CSI=1.

The CSI values of the seasonal flows for the three basin scales are shown in Figure 25. Results show that, except for the WRR1 product from Universiteit Utrecht (UniUt), the rest WRR products perform well, namely, CSI scoring above 0.8 for the three spatial scales. This indicates that although the models are demonstrated to exhibit strong biases of varied magnitude across the different models, in terms of the threshold scoring they all performed similarly well.

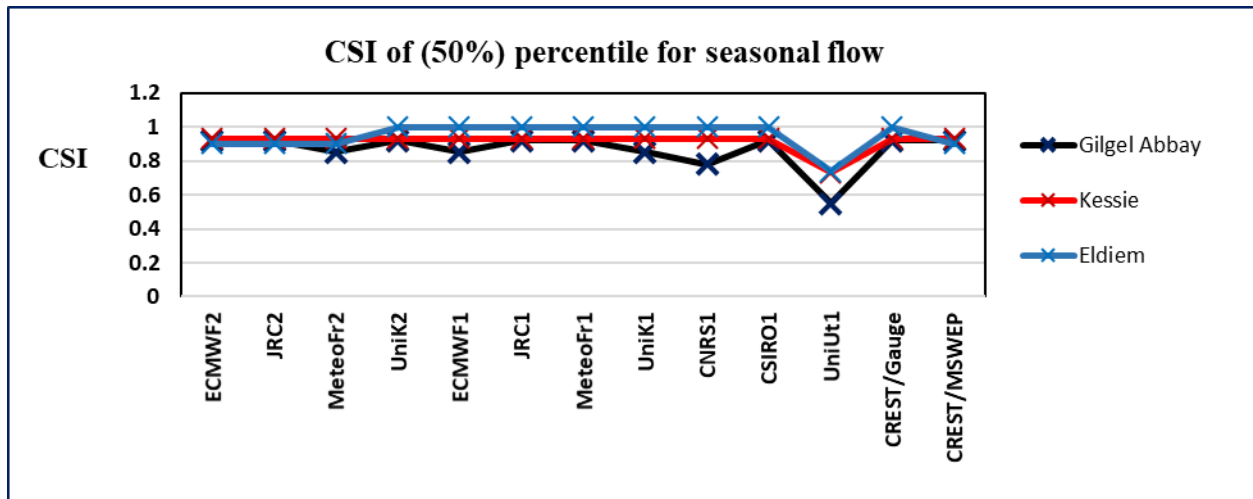


Figure 25. Critical Success Index (CSI) of the 50th-percentile seasonal flows of Gilgel Abbay, Kessie and Eldiem basin scales.

For the wet season (JJAS) monthly runoff we present the CSI for the 80th-percentile threshold. Results show basin scale dependence of the CSI metric for most of the models. The large-scale basin of Eldiem exhibits the highest CSI scores, followed by the medium-scale basin of Kessie and the weakest performance is for the small-scale basin of Gilgel Abbay as shown in Figure 26. Again, for the majority of the models we note similar performance across the different models, which highlights, regardless their performance differences, all models have good skill and the

ability to detect flow patterns indicating most or all models can be improved with local inputs data inputs. The global products performance show that the 80th monthly percentile is applicable mostly for large watersheds as shown in Figure 26. The large spatial scale of Eldiem shows better performance of the medium scale of Kessie. The medium scale of Kessie shows better performance of the small scale of Gilgel Abbay watershed. The WRR products perform poor, namely, CSI scoring below 0.4 for most of the products for small spatial scale in the upper Blue Nile basin. This indicates that although the models perform well for the wet seasonal flow of the 50th-percentile, but it shows the limitation to score well for the 80th-percentile for small scale of the basin. This indicate that it is potentially problematic to apply the global runoff products for management decisions in dry seasons for the small spatial scale of the basin.

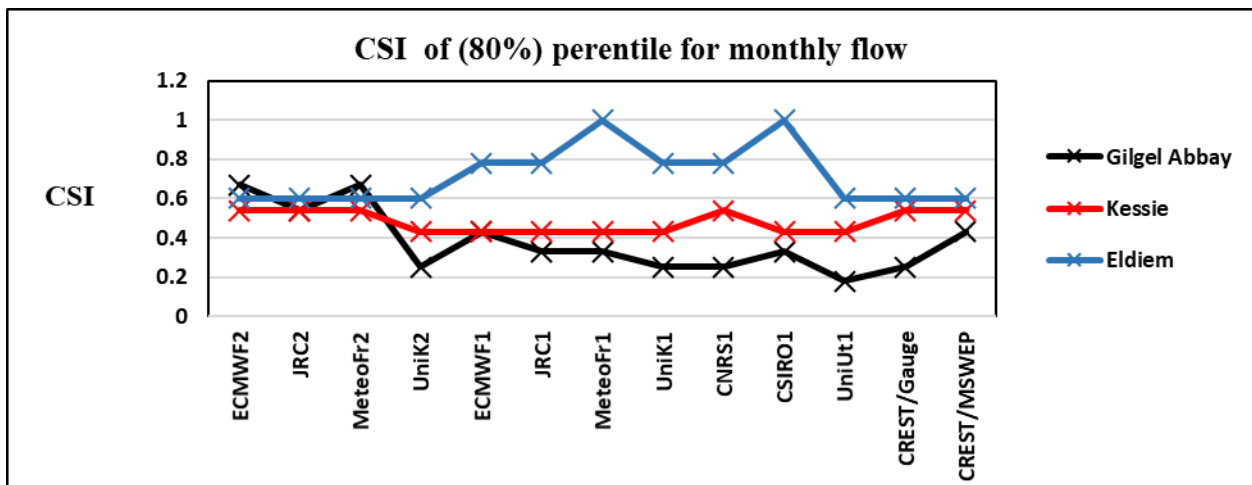


Figure 26. Critical Success Index (CSI) of the 80th-percentile monthly flow of Gilgel Abbay, Kessie and Eldiem basin scales.

The global product of UniK is validated for the other nine streamflow gauging stations in the upper Blue Nile basin. The validation is carried out from 1980 to 2012 for the nine stations shown in Table 12. The UniK product shows best estimates for the three case studies of Gilgel Abbay, Kessie

and Eldiem, however, the results of the other case studies show that the global product of UniK has the limitation to capture the observed flow for some gauging stations as shown in Table 12.

These all results caution the use of available global hydrological products without prior uncertainty evaluation, and the data needs bias correction or adjustment before for the application for planning, designing and management of water resources projects. The next chapter is carried out to improve the performance and uncertainty for the global product of UniK applying different dynamical bias correction for gauged and ungauged watersheds.

Table 12. Performance of global WRR2 (UniK) product for different case studies.

No.	Station		Uncorrected UniK
1	Main Beles (3,431 km ²)	NSCE	0.44
		Bias (100%)	-17.3
		CC	0.67
2	Chacha (41.8 km ²)	NSCE	0.12
		Bias (100%)	63.4
		CC	0.62
3	Guder (524 km ²)	NSCE	0.48
		Bias (100%)	21.7
		CC	0.77
		NSCE	0.19

4	Neshi (322 km ²)	Bias (100%)	-53.2
		CC	0.64
5	Didessa (Dembi) (1,806 km ²)	NSCE	0.30
		Bias (100%)	-38.3
		CC	0.62
6	Ribb (1,592 km ²)	NSCE	0.54
		Bias (100%)	8.6
		CC	0.80
7	Abbay (Bahir Dar) (15,321 km ²)	NSCE	-1.56
		Bias (100%)	107.7
		CC	0.61
8	Gumera (1,394 km ²)	NSCE	0.47
		Bias (100%)	15.4
		CC	0.78
9	Birr (978 km ²)	NSCE	0.34
		Bias (100%)	-36.4
		CC	0.63

5.4. Conclusive Remarks

In this work, we assess two versions of the earthH2Observewater resources reanalysis (WRR) river runoff products at three temporal resolutions (daily, monthly and seasonal). The earlier version WRR1 datasets are available at 0.5° spatial resolution and the newly released version (WRR2) is at 0.25° resolution. The WRR products are also compared to the performance of a locally-set distributed hydrological model forced with the global precipitation product of MSWEP. The study focused on three basin scales in the upper Blue Nile basin: Gilgel Abbay ($1,656 \text{ km}^2$), Kessie ($65,784 \text{ km}^2$), and Eldiem ($199,812 \text{ km}^2$) and daily, monthly and seasonal runoff. We evaluate model performances based on Q-Q plots and statistical metrics of NSCE, CC, RMSE, and Standard Deviation (S) determined for the period from 2000 to 2012. Our conclusions from the study are summarized below.

We found pronounced difference in the performance of WRR1 runoff products at all spatial and temporal scales considered in this study. No single product performs consistently better than the other products across the different spatio-temporal scales. Some WRR1 products perform better at the small basin Gilgel Abbay (e.g JRC product) while other products perform better at the larger basins at Kessie and Eldiem stations (e.g. CNRS product).

Unlike the coarser WRR1 runoff products, we found the high-resolution WRR2 runoff product from the Universitat Kassel provided a consistent and significantly improved runoff product across all spatial and temporal scales in the upper Blue Nile basin. This product stands out as a promising runoff dataset at range of spatial and temporal scales for water resources application in the upper Blue Nile basin. Other WRR products which provided better performance at different scales can equally be used at the specific temporal and spatial scale it performed well.

Comparison of the runoff product from the global hydrological models with the runoff simulated from a locally-set hydrological model (CREST) using global precipitation forcing of MSWEP, the latter appears to have a better edge than the earlier at daily and monthly temporal scales. However, at seasonal temporal scale, the (WRR2) Universitat Kassel product shows better performance than the locally simulated runoff. These models, which provide relatively higher performance at larger than monthly time scale, are potentially useful for assessment of long term regional climate and environmental trends.

On the basis of our application at a range of spatial and temporal scales in the nested watersheds of the upper Blue Nile basin, global water resources reanalysis (WRR) runoff products are improving steadily and can be used for water resources management with careful selection of the products. If robust locally set hydrological models are available, runoff simulated using global precipitation forcing provides comparable if not better results than runoff simulated from the best global hydrological model.

There is a desperate situation of data scarcity in Africa and a tendency of using any available data for planning, designing and management of water resources projects, the accuracy and reliability of global hydrological products vary from place to place and the use of these data products for actual projects locally without uncertainty evaluation will have associated risks and costs.

The next chapter is carried out to improve the uncertainty and performance of the global runoff product and to develop consistent set of gridded runoff product for water resources application for gauged and ungauged watersheds in data scarce region of the upper Blue Nile basin.

Chapter Six

6. Dynamical Bias Correction Procedure to Improve the WRR Gridded Daily Streamflow Data for Local Application in the Upper Blue Nile Basin

6.1. Introduction

Development of accurate hydrologic model predictions for current and potential future climatic conditions requires reliable spatially and temporally distributed climatic information [Hay and Clark, 2003; Jasper et al., 2002]. However, the availability and quality of climate data varies from region to region, and observations are sparse and irregularly distributed in many areas [El-Sadek et al., 2011; Haberlandt and Kite, 1998; Kanamaru and Kanamitsu, 2007]. Especially in Africa in which the rainfall gauging stations are sparse and irregularly distributed acquiring accurate and consistent hydro-meteorological data is challenging task. In view of this high-resolution satellite precipitation estimation in hydrological applications has been investigated for more than two decades [Guetter et al., 1996; Hossain and Anagnostou, 2004; Lakew et al., 2017; Mei et al., 2016; Nikolopoulos et al., 2010]. Along with high resolution satellite precipitation, recently reanalysis studies, that merge climate model predictions and observational data, have been implemented to develop long-term historical spatially distributed climatic variables such as precipitation, temperature, solar radiation, humidity and wind speed [Bromwich and Fogt, 2004; Overpeck et al., 2011].

Reanalysis products merge available observations with a state-of-the-art atmospheric (or more recently, coupled ocean–atmosphere–sea ice) model to derive the best estimate of the state of the atmosphere and land surface [Decker et al., 2012]. A typical reanalysis system consists of two main components, the forecast system and the data assimilation system. The role of the data

assimilation system is merging available observational data with the forecast model simulations [Seyyedi *et al.*, 2015b]. When global precipitation (satellite and reanalysis) products are used to force hydrologic models, they usually exhibit systematic and random errors and resolution effects in the simulation of river runoffs [M Bitew and Gebremichael, 2011; Lakew *et al.*, 2017; Seyyedi *et al.*, 2015b]. To improve the spatio-temporal availability of runoff data and avail consistent set of data for various global and regional water resources analysis, earthH2Observe generated global runoff WRR products. This study uses the runoff product developed by the Universitat Kassel (UniK) using global WaterGAP hydrological model. This runoff product is selected due to its superior performance shown in the evaluation analysis indicated in Chapter 5. The spatio-temporal dynamical bias correction procedure discussed have been used to improve the uncertainty imbedded in the global runoff product at local basin scale. The purpose of the study is to develop consistent set of gridded runoff product for water resources management for the gauged and ungauged watersheds.

6.2. Data and methods

6.2.1. Study Area

This study is conducted in the upper Blue Nile basin at twelve streamflow gauging stations. The gauging stations are Eldiem, Kessie, Gilgel Abbay, Beles, Chacha, Guder, Neshi, Didessa, Ribb Abbay (Bahir Dar), Gumea and Birr as shown in Figure 27.

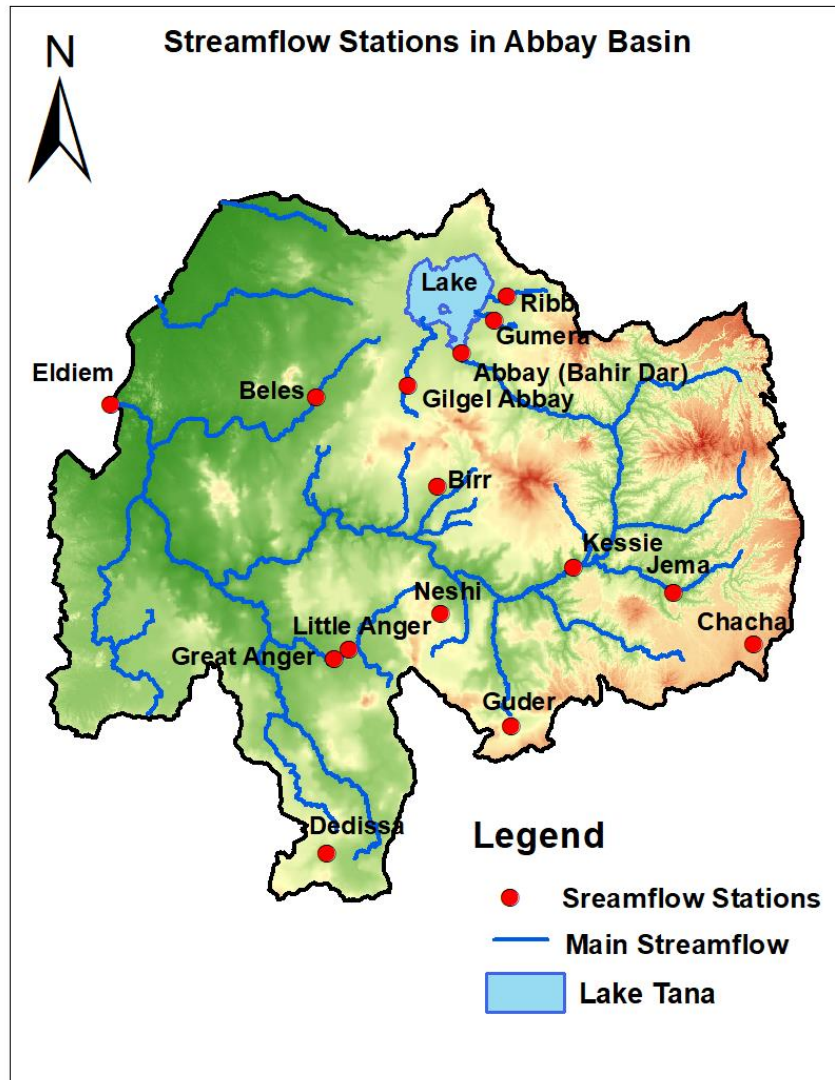


Figure 27. Location of streamflow stations in the upper Blue Nile basin.

6.2.2. Data Sets

The station river runoff dataset obtained from the Ethiopian Ministry of Water, Irrigation and Electricity used to evaluate the global grid-based reanalysis runoff data and serve as reference to estimate the bias factor of the corresponding pixel of the gridded global data and for the simulation. The gauging station network in Abbay basin is sparse and stations are unevenly distributed over the basin. From the whole streamflow gauging stations in Abbay basin, twelve streamflow gauging

stations that would cover the entire spatial coverage of the basin with better quality and long records from 1980 to 2010 are chosen for the dynamical bias correction and validation. The gauging stations are Eldiem, Kessie, Gilgel Abbay, Beles, Chacha, Guder, Neshi, Didessa, Ribb and Abbay (Bahir Dar). Two stations of Gumera and Birr used for the validation of the bias corrected new set of gridded river runoff data and shown in Figure 27. The updated streamflow data beyond 2010 for the twelve gauging stations is not available and the bias correction beyond 2010 is not carried out.

Two multi-model water resources reanalysis (WRR) runoff product of Universitat Kassel are available at 0.5° (WRR1) and 0.25° (WRR2) grid resolution, which was recently produced in the framework of a European Union project (earth2Observe). In the current study, the high-resolution (0.25°) of Universitat Kassel (WRR2) that has better performance from WRR products (see chapter 5) is considered for the bias correction analysis. The bias in the global runoff products should be corrected applying different dynamical bias correction schemes before the global runoff products are applied for water resources applications for local case studies.

Universitat Kassel Reanalysis Global Runoff Product

The University of Kassel (UniK, WaterGAP) is a global model of water availability and water use, has been developed to assess the current water resources situation and to estimate the impact of global change on the problem of water scarcity [Alcamo J, 2000; Doell *et al.*, 2001]. The WaterGAP model was developed at the Centre for Environmental Systems Research of the University of Kassel, Germany, in cooperation with the National Institute of Public Health and the Environment of the Netherlands (RIVM). The goals of the model are:

- to enable a comparison of the “freshwater situation” in different parts of the world, i.e. the uses and availability of freshwater to meet various objectives related to the requirements of society and aquatic ecosystems;
- to provide a long-term perspective (at least a few decades) on changes in global water resources.

The WaterGAP 2 model consists of two main components a Global Water Use model and a Global Hydrology model. The Water Use model takes into account basic socio-economic factors that lead to domestic, industrial and agricultural water use, while the Hydrology model incorporates physical and climate factors that lead to runoff and groundwater recharge [Alcamo *et al.*, 2003]. Two multi-model water resources reanalysis (WRR) runoff product of Universitat Kassel are available at 0.5° (WRR1) and 0.25° (WRR2) grid resolution with daily temporal resolution, which was recently produced in the framework of a European Union project (earthH2Observe). As introduced in the above section, this runoff product is used to improve the uncertainty imbedded in the global runoff product to produce the consistent set of gridded runoff data set for the Blue Nile basin using bias correction analysis.

6.2.3. Scheme for Bias Correction

Both global precipitation products of satellite and reanalysis rainfall estimates exhibit large systematic and random errors. Hence, the water resources reanalysis (WRR) global products consist of errors cascaded from precipitation forcing, global hydrological model formulation and processing simplification. Thus, bias in global runoff products should be evaluated and corrected before the global runoff products can be applied for water resources applications for local case studies. We use a multiplicative daily bias reduction estimator (Bias Factor, BF) to correct the bias for the global water resources reanalysis product studied in this research. For the analysis three

schemes of bias correction are applied and evaluated. These are the i) the temporal spatial variable bias factor, ii) temporal spatial constant bias factor and iii) spatial variable bias factors. The following section provides the three methods of bias correction schemes.

I. Temporal Spatial Variable, BF_{TSV}

In the current study, we estimate and correct the bias in Universitat Kassel (UniK) estimates as follows. The multiplicative daily bias factor (BF) at a certain UniK grid runoff data with an observed runoff can be formulated as follows.

$$BF_{TSV} = \frac{UniK(i, t)}{G(i, t)} \quad (8)$$

where G and UniK represent daily gauge and global Universitat Kassel runoff product, respectively, i refers to gauge location and t refers to a Julian day number. The subscript “TSV” stands for “Temporal Spatial Variable” since the bias factor in this formulation is estimated for a specific pixel with 0.25° spatial resolution and for a particular day. Each pixel has the corresponding bias factor value for a specific day and the bias factor is dynamic spatially and temporally. The daily bias factor is taken to be noisy to estimate the bias factor for the specific day of the basin. To minimize the daily noise effect and to produce stable bias factor, the monthly average bias factors are carried out at a monthly temporal scale. The result shows that the monthly and daily temporal scale bias factors are near similar, for the cause that the magnitude of the daily flow doesn’t change a lot throughout a month. But, in the monthly estimate of the bias factor, the monthly average flows (aggregated from daily) have extra missed values (not a number value, NaN) than the daily flow. Since there is no high variability of the daily bias factor (flow magnitude) throughout a month, and to minimize the monthly missed (NaN) values propagation to the neighbor pixels during interpolation, the daily temporal scale is applied for bias correction.

This scheme helps in adjusting the bias at a pixel based at 0.25° spatial and at a daily temporal scales (i.e., time and space varying), and is based on using the BF_{TSV} factor estimated from Equation (8). To apply a correction that accounts for spatial and temporal variability in the UniK bias, the pixel-based daily BF_{TSV} factors are spatially interpolated with the same UniK spatial 0.25° grid resolution. For the case of runoff estimates we use nearest method of spatial interpolation and it is applicable for the insertion of the bias factor values at each pixel for the entire basin with 0.25° spatial resolution to set the new corrected global runoff dataset. The interpolation yields a spatial and temporally varying field of BF that cover the entire study area. The Universitat Kassel daily runoff fields are then divided by the BF_{TSV} bias fields for the respective time windows to result in a new set of Universitat Kassel estimates that as such are bias-corrected in a temporally and spatially varying scheme.

II. Temporal Spatial Constant, BF_{TSC}

Temporal and spatial constant (TSC) bias correction: in this formulation the bias is obtained by UniK estimates over the entire spatial coverage and over the total duration of the sample.

$$BF_{TSC} = \frac{\sum_{t=1}^T \sum_{n=1}^N UniK(i, t)}{\sum_{t=1}^T \sum_{n=1}^N G(i, t)} \quad (9)$$

where n is the total number of gauges within the entire spatial coverage of the study and T is the full duration of the study period. The bias correction in this case is applied by dividing the UniK estimates by the bias factor, BF_{TSC} , to result in a new set of UniK estimates that are bias-corrected in a spatially and temporally-lumped scheme.

III. Spatial Variable, BF_{sv}

In this formulation the BF is temporally lumped over the entire domain but is still estimated for each gauge Equation (10):

$$BF_{SV} = \frac{\sum_{t=1}^T UniK(i, t)}{\sum_{t=1}^T G(i, t)} \quad (10)$$

The bias correction in this case is applied by dividing each UniK field by the bias factor, BF_{SV} , to result a new set of UniK estimates that are bias-corrected in a temporally lumped but spatially varying scheme.

The bias factor of the three schemes is computed by dividing the uncorrected UniK to the observed (gauged) runoff. This might lead unexpected extremely large or very small bias factor values. To avoid the propagation of extremely large and very small value of the bias factor to the neighbor pixels, while the bias factor spatial interpolation is being carried out, we restrict the bias factor value between 0.5 and 2. Since, the preliminary study of the global product of UniK bias value of most of the stations is between 0.5 and 2. The nearest method is applicable because the method doesn't exceed the minimum and maximum point values and the interpolated surface rarely passes through the sample points.

6.3. Result and Discussion

The result section dwells on the cascaded analysis and development of the new consistent 30 years gridded runoff dataset for the entire Blue Nile basin. The following analysis steps are implemented.

- i) Evaluation and selection of bias correction scheme (out of the three methods explained above)
- ii) Estimation of bias factor for the selected method using other set of 10 watershed runoff.
- iii) Spatial interpolation of the bias factor (BF) values with the global data resolution (0.25°) over the entire Blue Nile. The daily bias factors are spatially interpolated with the same UniK spatial 0.25° grid resolution. For runoff estimates we use nearest method

of spatial interpolation and it is applicable for the insertion of the bias factor values at each pixel for the entire basin with 0.25° spatial resolution to set the new corrected global runoff dataset. The interpolation yields a spatial and temporally varying field of BF that cover the entire study area.

- iv) Updating of the original UniK runoff product by dividing the spatially interpolated bias factor obtained in No. iii. The uncorrected (UniK) daily runoff fields are divided by the bias fields for the respective time windows to result a new set of bias corrected runoff product.

6.3.1. Evaluation of Bias Correction Schemes

To evaluate the three bias correction schemes that are applied for the gridded based global UniK runoff product, for the demonstration we select three streamflow gauging stations from the total chosen twelve stations. The three stations that are used in the evaluation in Chapters 4 and 5 representing three spatial scales, small (Gilgel Abbay, 1,656 km²), medium (Kessie, 65,784 km²) and large scale (Eldiem, 199,812 km²) are used to compare the bias correction schemes. The three (NSCE, Bias and CC) statistical comparison methods described in Section 2.9 are used to evaluate and select best the bias correction schemes. The result of the three statistical metrics shows that the bias corrected data, applying the bias factor of temporal, spatial variable (BF_{TSV}) outperforms from the other method of bias correction for the three different spatial scales as shown in Table 13. Eldiem, the large basin scale, the NSCE performance shows high improvement from 0.80 to 0.95 for the uncorrected and corrected global river runoff product respectively applying BF_{TSV} bias correction scheme. The bias also shows high reduction from 13.9 % to 1.2 % as the UniK product is corrected using BF_{TSV}. The same with the large basin scale of Eldiem, the medium scale of Kessie shows high improvement of NSCE and reduction of bias applying BF_{TSV} scheme. The

NSCE improves from 0.66 to 0.93 and bias reduces from -12% to 7.2% for the BF_{TSV} scheme of bias correction for medium scale of Kessie. The NSCE performance result from Gilgel Abbay of small basin scale makes certain that the scheme of bias correction using BF_{STV} is applicable for the three case studies of small, medium and large basin scales based on the NSCE performance. The bias efficiency metric shows that the result from BF_{sv} scores minimum bias value for medium and large basin scales, however, the bias value of Gilgel Abbay shows highest from the other schemes. The result of bias of Gilgel Abbay and the performance of NSCE of the three case studies shows that the scheme from BF_{STV} is applicable for the runoff bias correction in the upper Blue Nile basin.

Table 13. Different schemes of bias correction for the three-different sized watersheds from 2000 to 2010.

Station		Uncorrected UniK	Correction Schemes for UniK		
			BF _{TSV}	BF _{TSC}	BF _{sv}
Eldiem (199,812 km ²)	NSCE	0.80	0.95	0.76	0.83
	Bias (100%)	13.9	1.2	21.3	0.1
	CC	0.91	0.98	0.91	0.91
Kessie (65,784km ²)	NSCE	0.66	0.93	0.60	0.58
	Bias (100%)	-12	-7.2	-3.9	0.38
	CC	0.83	0.97	0.81	0.81
Gilgel Abbay (1,656 km ²)	NSCE	0.66	0.76	0.67	0.61
	Bias (100%)	-21.7	-5.3	-16.7	-31.3
	CC	0.83	0.87	0.83	0.83

6.3.2. Evaluation of the Selected Bias Correction Scheme and Runoff

The result from the evaluation of the three bias correction schemes of the recent period from 2000 to 2010 shows the BF_{TSV} has high improvement and is more applicable for the upper Blue Nile basin. Considering this we implement this scheme of bias correction applying the twelve gauging stations of Eldiem, Kessie, Gilgel Abbay, Beles, Chacha, Guder, Neshi, Didessa, Ribb, Abbay (Bahir Dar), Gumera and Birr in the upper Blue Nile basin for the long period from 1980 to 2010 at daily temporal scale. The stations listed from 1 to 10 are used for the bias factor estimation in the upper Blue Nile basin (Table 14). The station from Gumera and Birr are not utilized for the bias factor estimation and for bias adjustment, these stations are only used to validate the new set of bias corrected product of UniK.

The bias factor of the 10 watersheds is interpolated over the basin at the same grid resolution as the product. The gridded uncorrected UniK runoff product is divided by the interpolated bias factor at each pixel. The new corrected runoff product is evaluated over the 12 watersheds as shown in Table 14.

The results show that the bias correction applying BF_{TSV} for the gridded runoff data of UniK shows high improvement for long-term data for the 12 watersheds. Nash-Sutcliffe coefficient (NSCE) improves by 55 % and bias decreases by 22% for twelve streamflow gauging stations (Table 14). The result from the stations used for validation shows that the bias corrected data applying the BF_{TSV} scheme for the long-term data from 1980 to 2010 shows improvement and applicable for the upper Blue Nile basin. The result from Gumera shows high improvement and Birr also shows encouraging improvement. Gumera and Birr NSCE performance improves by 23% and 12%, the bias also reduces by 15% and 11% for the two stations respectively.

The uncorrected UniK product of NSCE performance for long-term data from 1980 to 2010 shows that only Eldiem shows better skill above 0.70 and the rest shows relatively less performance below 0.6 (except Gilgel Abbay). Especially, Abbay (Bair Dar) shows very poor performance of -1.56 that is due to the effect of Lake Tana, that shows the uncorrected product of UniK has less efficiency for station that has a reservoir effect at the upstream of the streamflow gauging station. Kessie station has NSCE value of 0.37 for long records from 1980 to 2010 (Table 14) and 0.66 for the recent year from 2000 to 2010 (Table 13) for the uncorrected UniK runoff. This indicates that the recently used rating curve to convert the staff reading to streamflow of Kessie is much better than the earlier rating curve to estimate the global product of the uncorrected of UniK. The bias evaluation of the uncorrected UniK product shows that the whole stations scores above 5% of bias. From the total stations except Kessie (7%) and Ribb (8.6%) the rest stations score above 15% of bias, especially Neshi (-53.2%), Chacha (63.4%) and Abbay at Bahir Dar (107.7%) score above 50% of bias value. The station from Abbay (Bahir Dar) shows the highest bias value and the poorest performance, this is attributable to the effect of the Lake Tana located at the upstream of the gauging stations. These all results show that the uncorrected global gridded river runoff data has limitation to capture the observed river runoff data in the upper Blue Nile basin and the uncorrected global data needs adjustment or bias correction before the data is used for various water resources applications. Especially, the uncorrected global data doesn't consider the reservoir effects and the routing effect of the reservoir outflows, these show that applying the raw global river runoff data need caution to use for water resources application without bias adjustment.

The bias corrected UniK runoff shows high NSCE efficiency improvement relative to the uncorrected UniK runoff for the long-term data from 1980 to 2010 as shown in Table 14. Of the total 12 stations that have been considered in this study, 33% of the stations score above 0.9 and

50% of the stations scores above 0.8 of NSCE in the corrected UniK new set of gridded river runoff product. Highest NSCE value is observed at Gilgel Abbay (0.96), Eldiem (0.95), Guder (0.94) and Ribb (0.93) that score above 0.93. The corrected UniK runoff NSCE values shows that Kessie and Guder show maximum increment of NSCE value from 0.37 to 0.83 and from 0.48 to 0.94 respectively. These all NSCE result show that the bias corrected new set of gridded river runoff data (UniK, corrected) capture very well most of the gauging stations with high improvement relative to the uncorrected UniK runoff product. Based on bias performance evaluation of the corrected gridded UniK runoff product, from the total twelve stations more than half of the stations scores less than 5% of bias, such as Ribb and Gumera score the minimum bias value of -1.5% and 1.9% respectively. The corrected new set of data shows that only Abbay at Bahir Dar and Neshi score maximum value of bias, 30.5% and -36.9%, respectively. More than 80% of the stations score below 25% bias value for the new set of bias corrected river runoff gridded data. These all results show that the bias adjusted new set of high resolution (Daily,0.25°) gridded river runoff data shows high performance to estimate the observed river runoff and can be in use as one source of streamflow data for the data scarce region of the upper Blue Nile basin, Ethiopia.

Table 14. Performance of the new set of bias corrected UniK (BF_{Tsv}) river discharge from 1980 to 2010.

No.	Station		Uncorrected UniK	Corrected UniK (BF _{Tsv})
	Eldiem	NSCE	0.73	0.95
		Bias (100%)	24.2	3.1

1	(199,812 km ²)	CC	0.90	1.00
2	Kessie (65,784km ²)	NSCE	0.37	0.83
		Bias (100%)	7	-3.7
		CC	0.66	0.91
3	Gilgel Abbay (1,656 km ²)	NSCE	0.67	0.96
		Bias (100%)	-18	-4.4
		CC	0.83	0.98
4	Main Beles (3,431 km ²)	NSCE	0.44	0.80
		Bias (100%)	-17.3	-9.5
		CC	0.67	0.91
5	Chacha (41.8 km ²)	NSCE	0.12	0.42
		Bias (100%)	63.4	23.1
		CC	0.62	0.73
6	Guder (524 km ²)	NSCE	0.48	0.94
		Bias (100%)	21.7	5.9
		CC	0.77	0.97
7	Neshi (322 km ²)	NSCE	0.19	0.54
		Bias (100%)	-53.2	-36.9
		CC	0.64	0.83
8	Didessa (Dembi) (1,806 km ²)	NSCE	0.30	0.74
		Bias (100%)	-38.3	-20.2
		CC	0.62	0.89
		NSCE	0.54	0.93

9	Ribb (1,592 km ²)	Bias (100%)	8.6	-1.5
		CC	0.80	0.97
10	Abbay (Bahir Dar) (15,321 km ²)	NSCE	-1.56	0.68
		Bias (100%)	107.7	30.5
		CC	0.61	0.89
11	Gumera (1,394 km ²)	NSCE	0.47	0.70
		Bias (100%)	15.4	1.9
		CC	0.78	0.84
12	Birr (978 km ²)	NSCE	0.34	0.46
		Bias (100%)	-36.4	-25.6
		CC	0.63	0.71

Quantile-Quantile (Q-Q) plot

Figure 28 shows the Q-Q plot of the twelve stations of both uncorrected and bias corrected of the global UniK gridded runoff products with respect to the observed streamflow (reference). The uncorrected product shows high deviation from the reference of the 45-degree line. From the total stations of the case studies the uncorrected UniK runoff product shows that only Gumera, Ribb and Guder stations show good agreement with the reference line. Eldiem and Kessie also shows modest skill for the uncorrected river runoff product, the rest stations show poor performance especially Birr, Didessa and Neshi show very poor presentation to capture high quantiles. This indicates that the uncorrected river runoff has the limitation to capture the quantile of the observed streamflow in the upper Blue Nile basin and needs bias adjustment for the application of water management.

The bias corrected river runoff gridded global high-resolution product shows high improvement for all of the stations as shown in Figure 28. Especially, 50% of the overall stations such as Eldiem, Kessie, Gilgel Abbay, Guder, Ribb and Gumera show high improvement and performance to capture both low and high flow quantile estimates. Stations of Beles, Chacha and Birr show relatively modest skill to capture the high flow quantiles for the corrected UniK river runoff product. The same with the station from Abbay at Bahir Dar that has the effect of Lake Tana, the station from south-west part of the basin, Didessa and Neshi show improvement from the uncorrected global product but, still the corrected river runoff product doesn't capture very well the reference line. This indicates that the global product of uncorrected UniK has high bias value to estimate the observed river runoff for the stations that have upstream reservoir effect and the stations that are located at the south-west of the basin. The inaccuracy and inconsistency of the observed streamflow data has high impact on the evaluation of the product discussed. The whole Q-Q plot manifests that the bias corrected global high-resolution gridded river runoff product shows high improvement to estimate the quantiles of the observed river runoff compared with the uncorrected global product. On the basis of this exhaustive analysis and validation study of the dynamical bias correction approach, we produced 30 years (1980-2010) consistent set of bias corrected gridded (0.25°) daily runoff products for data as one source of data for planning, design and management water resources application for gauged and ungauged watersheds in the upper Blue Nile basin, Ethiopia.

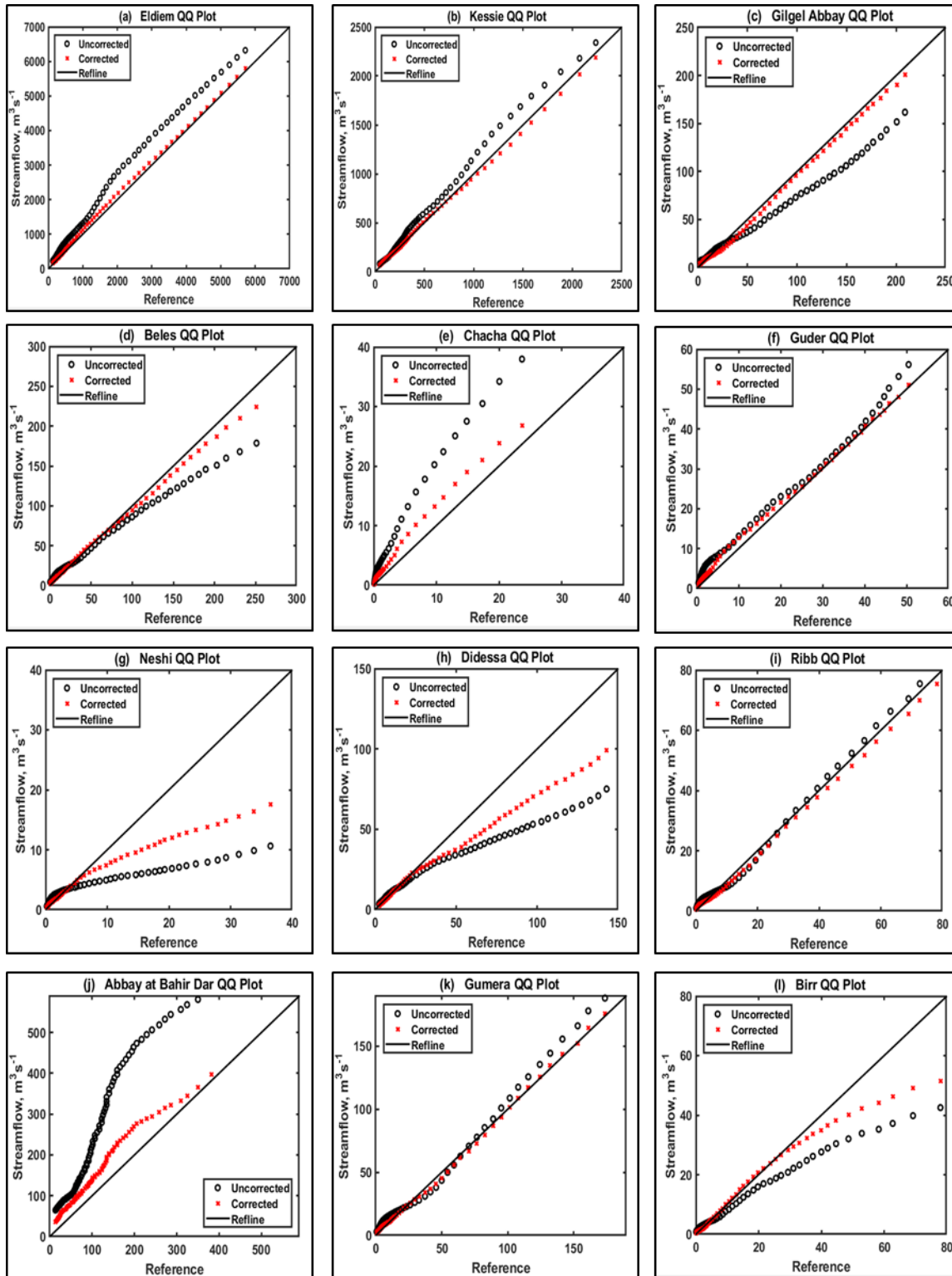


Figure 28. QQ plots of the corrected and uncorrected UniK products for different stations.

6.4. Conclusive Remarks

Evaluation of global data products including reanalysis river runoff has been carried out to estimate the observed streamflow for local case studies and the results show that the global products from both precipitation and reanalysis river runoff have the limitation to capture the observed streamflow. Still, not a lot has been done to demonstrate how these products, especially the global river runoff product can be ready for various water resources application by reducing their errors. This research focuses on improving the quality and availability of local river runoff data applying three dynamical bias correction schemes, by means of globally accessible water resources reanalysis river runoff product (WRR) of Universitat Kassel available at high resolution (Daily, 0.25°) for data scarce region of the upper Blue Nile basin, Ethiopia. Twelve streamflow stations are considered for bias correction and for validation of the bias corrected product using long term historical data from 1980 to 2010 at daily temporal scales. The outcomes contribute to efforts that aim towards enhancing the real-world applicability of reanalysis river runoff product.

The uncorrected reanalysis river runoff product of Universitat Kassel (UniK) has high bias range from 7% to 100.7% and the NSCE performance is between -1.56 to 0.7 for the streamflow gauging stations in the upper Blue Nile basin. This indicates that the uncorrected global UniK river runoff product has error river runoff generation mechanisms to estimate the observed streamflow for the entire basin river network and needs bias correction or adjustment for water resources applications.

For the analysis we apply three bias correction schemes to correct the bias UniK global reanalysis river runoff estimates. Temporal spatial variable, temporal spatial constant and spatial variable bias factors are valued to correct UniK estimation. The bias corrected data, applying the bias factor of temporal, spatial variable (BF_{TSV}) outperforms from the other method of bias correction for

different spatial scales, indicates that the UniK river runoff temporal variation as accounting for it significantly reduced the runoff bias.

The result shows that the bias correction applying BF_{TSV} for the gridded runoff data of UniK shows high improvement for long term data for the case studies. Nash-Sutcliffe coefficient (NSCE) improves by 55 % and bias decreases by 22% for streamflow gauging stations that are used for the bias factor estimation and validation in the upper Blue Nile basin. The stations used for validation shows that the bias corrected data applying the BF_{TSV} scheme for the long records from 1980 to 2010 shows improvement and the scheme is applicable for the upper Blue Nile basin. From the total stations, 33% of the stations score above 0.9 and 50% of the stations scores above 0.8 of NSCE and more than half of the stations scores less than 5% of bias in the corrected UniK new set of gridded river runoff product.

These all results show that the bias adjusted new set of high resolution (0.25°) gridded river runoff daily product can be used as one source of streamflow data for both gauged and ungauged watersheds in the data scare region of the upper Blue Nile basin, Ethiopia. The bias correction is applied only using ten, and validation uses two streamflow gauging stations in the upper Blue Nile basin. Further investigation, using a reasonably large number of stations with high-quality and recently updated streamflow data may be needed.

Chapter Seven

7. Summary, Conclusions and Recommendations

7.1. Summary and Conclusions

Water resource applications in regions without good quality and reliable temporal rainfall data are complicated by a lack of adequate spatial coverage of datasets. Hydro-meteorological datasets have significant importance on the sustainable water resource management and applications taking in to account the datasets as hydrological model inputs and spatiotemporal distributed water over the landscape as model output. However, acquiring accurate and consistent hydro-meteorological data is challenging task throughout the world specially in Africa in which the rainfall gauging stations are limited and sparsely located. For example, in many African countries, the spatial coverage of hydrometeorological stations is very low. An African Climate Policy center (ACPC) assessment report [(ACPC), 2011] indicates while the spatial coverage of African climate stations is in the order of 1 station per 27,347 km², it is 1 station per 1244 km² in Germany. Similarly, the coverage of the streamflow stations in Africa is very low. There are 888 gauging stations in Africa in an area of 21,300,000 km² (without including the area of the Sahara), while there are 1150 streamflow gauges per 357,114 km² area in Germany. The situation in Ethiopia and upper Blue Nile is not different from this global picture. This contrasting revelation indicates the importance of supplementing the continent's hydro-information system with remotely sensed satellite and reanalysis products. Therefore, high-resolution precipitation and reanalysis runoff products have major role in hydrological analysis and water resource application in such data limited regions.

It is well known that satellite and reanalysis rainfall values are estimates that are subject to a variety of error sources and these high-resolution satellite and reanalysis products have their own errors. Using precipitation data from satellites or global reanalysis products to force hydrologic models,

exhibits complex rainfall systematic and random errors and propagates in the simulation of river runoffs. Alternatively, water resources reanalysis (WRR) river runoff, global products are emerging with high resolution applying global models that merge observations and models to provide long and consistent global runoff data. The global runoff products have limitations to capture the quantity and represent amply the hydrograph dynamics runoff of the corresponding station observed runoff of local case studies and the global datasets require validation at local basin scale. Based on these, different studies as part of this dissertation have been conducted for the evaluation and to improve the performance of the global datasets to estimate the river runoff.

1. The first study focuses on the hydrological performance of various satellite and reanalysis global gridded precipitation products simulating the daily flow time for Gilgel Abbay watershed. The Coupled Routing and Excess Storage (CREST) fully distributed hydrological model [Xinyi Shen and Hong, 2014] is used for this evaluation and comparison of study. The performance evaluation of various satellite and reanalysis precipitation datasets for water resource applications in the upper Blue Nile basin, Ethiopia, applying different calibration and simulation conditions (modes) to improve the performance of the precipitation products to simulate the river runoff and the results indicate that;

- Examination of the parameters of the model indicates Saturated Hydraulic Conductivity (Ksat), Mean Water holding capacity (WM) and the overland runoff velocity coefficient (coeM) show wide range parameter values for the different precipitation products of independent calibration. These three parameters are less sensitive relative to the other parameters. It is also noted that other model

parameters are sensitive and remain nearly the same for the tested precipitation products. The most sensitive parameter is RainFact.

- In terms of comparing the performance of the satellite and reanalysis precipitation products, all products under independent and RainFact calibration mode reproduce daily streamflow better than the simulation mode with NSCE performance above 70% in calibration and 60% in validation periods.
- The results indicate the precipitation products consistently provide a better performance of runoff estimation when they are independently calibrated and under RainFact calibration than simulation mode of the products. We conclude as long as each product is calibrated independently or under RainFact calibration, global precipitation products can provide enough information for water resource management in data-scarce regions of upper Blue Nile basin.
- Independent calibration is time consuming and demands several times for calibration and it is exclusively applicable for small spatial scale watersheds. For larger basin scales in the upper Blue Nile basin, considering various precipitation products and carrying out independent calibration of specific product using fully distributed hydrological model CREST is much time taking. For this reason, it is recommended to carry out RainFact calibration mode for each precipitation product which presents an equivalent performance with the independent calibration. The next chapter considers different sized watersheds (small, medium and large basin scales) and several precipitation products. The evaluation of the products is made applying the RainFact calibration and simulation modes.

2. This study compares hydrological performance of globally available satellite and reanalysis precipitation products in data scarce basins of Africa. The study is conducted in the upper Blue Nile basin over multi-scales (1,656–199,812 km²) focusing on multi-year (2000–2012) daily simulation and analysis. Grid based, fully distributed hydrological model Coupled Routing and Excess Storage (CREST) is calibrated using in-situ observed rainfall data. The comparison is performed in two modes. In the first mode, the gauge calibrated CREST model is used to simulate and compare the performance of each group of products. In the second mode, the dominant model parameter which controls the volume of precipitation products is tweaked (or calibrated) for each product and compared. The evaluation compares two satellite precipitation products of gauge adjusted Climate Prediction Center Morphing Technique (CMORPH) and Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis 3B42 version 7 (TMPA) and three reanalysis precipitation products of ERA-Interim (ERA-Interim), Global Precipitation Climatology Centre (GPCP) and Multi-Source Weighted Ensemble Precipitation (MSWEP) and the results indicate;

- Overall, except the ERA-Interim reanalysis precipitation product, both the satellite and reanalysis precipitation products produced comparatively consistent performance for the different sized watersheds in the upper Blue Nile basin.
- Tweaking of the RainFact precipitation parameter further improved the performance of all products. The performance of the MSWEP outranks both the satellite and reanalysis categories of precipitation products in its runoff predictive ability.

- The MSWEP reanalysis precipitation product can be a priori alternative source of data (under fine-tuning) for various water resources applications in the upper Blue Nile basin for its predictive ability and the length of data availability which reanalysis products have advantage over the satellite products. This by no means undermines the potential benefits of the satellite and the reanalysis products.
3. The increasing availability of global observation datasets, both from in-situ and remote sensors, along with advancements in earth system models and data assimilation algorithms, have led to the generation of a number of water resources reanalysis products that are available at global scale and high spatial and temporal resolutions. These products hold a great potential for water resource applications, but their level of uncertainty needs to be evaluated at local scale. In this work, we evaluate the runoff product from two multi-model global water resources reanalysis (WRR) river runoff, available at 0.5° (WRR1, seven products) used WFDEI and 0.25° grid resolution (WRR2, four products used MSWEP), that are produced in the framework of a European Union project (earth2Observe) and the analysis indicates.
- The recently released WRR2 Universitat Kassel (UniK) product exhibits consistently better performance statistics than the earlier coarser resolution WRR1 and the rest of WRR2 products at all range of temporal and spatial scale evaluated.
 - Runoff simulated using locally set CREST hydrological model forced with Multi Source Weighted Ensemble Precipitation (MSWEP) global precipitation products have better edge than all other products including UniK for daily and monthly runoff simulations, for seasonal flow the UniK product shows better performance.

- Global hydrological products can be a data source for various water resources planning and management application in data scarce areas of Africa. The global runoff products are not ready to be used in sectoral design and management. This study cautions the use of available global hydrological products without prior uncertainty evaluation.
4. The global runoff products (WRR) have limitations to capture the quantity and represent amply the hydrograph dynamics runoff of the corresponding station observed runoff of local case studies and need correction and fine-tuning to use the product for water resources applications for watersheds with no or limited records. This research focuses on improving the uncertainty, performance and availability of global river runoff data applying dynamical bias correction. Water resources reanalysis river runoff product (WRR) of Universitat Kassel at high resolution (Daily, 0.25°) is used for the analysis of bias correction. Three schemes of dynamical bias correction are applied: temporal spatial variable, temporal spatial constant and spatial variable. The results are;
- The bias corrected data, applying the bias factor of temporal, spatial variable (BF_{TSV}) outperforms from the other method of bias correction for different spatial scales, indicates that the UniK river runoff temporal variation as accounting for it significantly reduced the runoff bias.
 - The new set of Universitat Kassel runoff product applying the temporal spatial variable scheme shows high improvement. Nash-Sutcliffe coefficient (NSCE) improves by 55 % and bias decreases by 22% for twelve river runoff gauging stations in the upper Blue Nile basin for long-term from 1980 to 2010.

- On the basis of this exhaustive analysis and validation study of the dynamical bias correction approach, we produced 30 years (1980-2010) consistent set of bias corrected gridded (0.25°) daily runoff products for data as one source of data for planning, design and management water resources application for gauged and ungauged watersheds in the upper Blue Nile basin, Ethiopia.

7.2. Recommendations

Satellite and reanalysis precipitation products consistently provide a better performance of runoff estimation when they are independently calibrated than simulation modes of the products and as long as each satellite and reanalysis product is calibrated independently, global precipitation products can provide enough information for water resource management in data-scarce regions of upper Blue Nile basin.

Independent calibration of the specific precipitation products for large spatial scales of the basin applying the grid based fully hydrological model (CREST) is time consuming. Tweaking of the RainFact precipitation parameter further improved the performance of all products, less time consuming and applicable for large temporal scales in the upper Blue Nile basin.

The tweaking of RainFact parameter for the reanalysis precipitation product of MSWEP outranks both the satellite and reanalysis categories of precipitation products in its runoff predictive ability. The MSWEP reanalysis precipitation product (under fine-tuning) can be a priori alternative source of data for various water resources applications in the upper Blue Nile basin for its predictive ability and the length of data availability which reanalysis products have advantage over the satellite products. This by no means undermines the potential benefits of the satellite and the reanalysis products.

Runoff simulated using locally set CREST hydrological model forced with Multi Source Weighted Ensemble Precipitation (MSWEP) global precipitation products have better edge than all other global water resources reanalysis (WRR) products including UniK for daily and monthly temporal scales, but the UniK has better performance for seasonal temporal resolution. Global hydrological products can be a data source for various water resources planning and management application in data scarce areas of Africa. This study cautions the use of available global hydrological products without prior uncertainty evaluation and the use of these data products for actual projects locally without uncertainty evaluation will have associated risks and costs.

The global runoff products (WRR) have limitations to capture the quantity and represent amply the hydrograph dynamics runoff of the corresponding station observed runoff of local case studies and need correction and fine-tuning to use the product for water resources applications for ungauged and gauged watersheds with limited records. On the basis of this exhaustive analysis and validation study of the dynamical bias correction approach, we produced 30 years (1980-2010) consistent set of bias corrected gridded (0.25°) daily runoff products for data as one source of data for planning, design and management water resources application for gauged and ungauged watersheds in the upper Blue Nile basin, Ethiopia.

The bias correction and validation are applied only using ten and two streamflow gauging stations respectively in the upper Blue Nile basin. We recommend further investigation with uncertainty analysis based on a reasonably large number of stations with high-quality and recently updated streamflow data for the global runoff products.

References

- (ACPC), A. C. P. C. (2011), Climate Change and Water in Africa: Analysis of Knowledge Gaps and Needs, *African Climate Policy Centre: Addis Ababa, Ethiopia*.
- Alcamo, J., P. DÖLl, T. Henrichs, F. Kaspar, B. Lehner, T. RÖSch, and S. Siebert (2003), Development and testing of the WaterGAP 2 global model of water use and availability, *Hydrological Sciences Journal*, 48(3), 317-337.
- Alcamo J, H. T., Rösch T (2000), Global modeling and scenario analysis. In:Rijsberman F (ed): World Water Scenarios – Analyses., *Earthscan Publications, London*.
- Allen, R., L. Pereira, D. Raes, and M. Smith (1998), Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56, *FAO - Food and Agriculture Organization of the United Nations, Rome*.
- Amir, A., M. Ali, N. Hamidreza, and B. Ali (2012), Systematic and random error components in satellite precipitation data sets, *Geophysical Research Letters*, 39(9).
- Andersson, E. (2015), User Guide to ECMWF Forecast Products edited, UK.
- Balsamo, G., A. Beljaars, K. Scipal, P. Viterbo, B. van den Hurk, M. Hirschi, and A. K. Betts (2009), A Revised Hydrology for the ECMWF Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System, *Journal of Hydrometeorology*, 10(3), 623-643.
- Balsamo, G., et al. (2015), ERA-Interim/Land: a global land surface reanalysis data set, *Hydrol. Earth Syst. Sci.*, 19(1), 389-407.
- Beck, H., A. van Dijk, V. Levizzani, J. Schellekens, D. Miralles, B. Martens, and A. Roo (2017), *MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data*, 589-615 pp.

Beck, H. E., A. I. J. M. van Dijk, A. de Roo, E. Dutra, G. Fink, R. Orth, and J. Schellekens (2017), Global evaluation of runoff from 10 state-of-the-art hydrological models, *Hydrol. Earth Syst. Sci.*, *21*(6), 2881-2903.

Bitew, M., and M. Gebremichael (2011), *Assessment of high-resolution satellite rainfall for streamflow simulation in medium watersheds of the East African highlands*, 1147-1155 pp.

Bitew, M. M., M. Gebremichael, L. T. Ghebremichael, and Y. A. Bayissa (2012), Evaluation of High-Resolution Satellite Rainfall Products through Streamflow Simulation in a Hydrological Modeling of a Small Mountainous Watershed in Ethiopia, *Journal of Hydrometeorology*, *13*(1), 338-350.

Bland, J. M., and D. G. Altman (1996), Statistics Notes: Measurement error, *BMJ*, *313*(7059), 744.

Bromwich, D. H., and R. L. Fogt (2004), Strong Trends in the Skill of the ERA-40 and NCEP–NCAR Reanalyses in the High and Midlatitudes of the Southern Hemisphere, 1958–2001, *Journal of Climate*, *17*(23), 4603-4619.

Burek, P., A. Roo, and J. van der Knijff (2013), *LISFLOOD - Distributed Water Balance and Flood Simulation Model - Revised User Manual*.

Chu, W., X. Gao, and S. Sorooshian (2010), *Improving the Shuffled Complex Evolution Scheme for Optimization of Complex Nonlinear Hydrological Systems: Application to the Calibration of the Sacramento Soil-Moisture Accounting Model*.

Dantec-Nédélec, S., et al. (2017), Testing the capability of ORCHIDEE land surface model to simulate Arctic ecosystems: Sensitivity analysis and site-level model calibration, *Journal of Advances in Modeling Earth Systems*, *9*(2), 1212-1230.

Decker, M., M. A. Brunke, Z. Wang, K. Sakaguchi, X. Zeng, and M. G. Bosilovich (2012), Evaluation of the Reanalysis Products from GSFC, NCEP, and ECMWF Using Flux Tower Observations, *Journal of Climate*, 25(6), 1916-1944.

Dee, D. P., et al. (2011), The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Quarterly Journal of the Royal Meteorological Society*, 137(656), 553-597.

Dinku, T., P. Ceccato, E. Grover-Kopec, M. Lemma, S. J. Connor, and C. F. Ropelewski (2007), Validation of satellite rainfall products over East Africa's complex topography, *International Journal of Remote Sensing*, 28(7), 1503-1526.

Doell, P., J. Alcamo, T. Henrichs, F. Kaspar, B. Lehner, T. Rösch, and S. Siebert (2001), *The global integrated water model WaterGAP 2.1*.

Donaldson., R.J., R. M. Dyer., and M. Krauss. (1975), An objective evaluator of techniques for predicting severe weather events, *Preprints, Ninth Conf. on Severe Local Storms, Amer. Meteor. Soc., Norman, OK*, 321-326. .

Doswell, C., R. Davies-Jones, and D. L. Keller (1990), *On Summary Measures of Skill in Rare Event Forecasting Based on Contingency Tables*.

Duan, Q., S. Sorooshian, and V. Gupta (1992), Effective and Efficient Global Optimization for Conceptual Rainfall-Runoff Models *Water Resources Research*, 28.

Ducoudré, N. I., K. Laval, and A. Perrier (1993), SECHIBA, a New Set of Parameterizations of the Hydrologic Exchanges at the Land-Atmosphere Interface within the LMD Atmospheric General Circulation Model, *Journal of Climate*, 6(2), 248-273.

El-Sadek, A., M. Bleiweiss, M. Shukla, S. Guldan, and A. Fernald (2011), Alternative climate data sources for distributed hydrological modelling on a daily time step, *Hydrological Processes*, 25(10), 1542-1557.

Elshamy, M. E., I. A. Seierstad, and A. Sorteberg (2009), Impacts of climate change on Blue Nile flows using bias-corrected GCM scenarios, *Hydrol. Earth Syst. Sci.*, 13(5), 551-565.

Erkossa, T., S. Awlachew, A. Hailelassie, and A. Denekew Yilma (2009), *Impacts of improving water management of smallholder agriculture in the Upper Blue Nile Basin*.

Gerten, D., S. Rost, W. von Bloh, and W. Lucht (2008), Causes of change in 20th century global river discharge, *Geophysical Research Letters*, 35(20), n/a-n/a.

Gleick, P. H., and G. F. White (1993), *Water in Crises : A Guide to the World's Fresh Water Resources*, Oxford University Press.

Guetter, A. K., K. P. Georgakakos, and A. A. Tsonis (1996), Hydrologic applications of satellite data: 2. Flow simulation and soil water estimates, *Journal of Geophysical Research: Atmospheres*, 101(D21), 26527-26538.

Haberlandt, U., and G. W. Kite (1998), Estimation of daily space-time precipitation series for macroscale hydrological modelling, *Hydrological Processes*, 12(9), 1419-1432.

Habib, E., A. Haile, N. Sazib, Y. Zhang, and T. Rientjes (2014a), *Effect of Bias Correction of Satellite-Rainfall Estimates on Runoff Simulations at the Source of the Upper Blue Nile*, 6688-6708 pp.

Habib, E., A. Haile, N. Sazib, Y. Zhang, and T. Rientjes (2014b), Effect of Bias Correction of Satellite-Rainfall Estimates on Runoff Simulations at the Source of the Upper Blue Nile, *Remote Sensing*, 6(7), 6688.

Hay, L. E., and M. P. Clark (2003), Use of statistically and dynamically downscaled atmospheric model output for hydrologic simulations in three mountainous basins in the western United States, *Journal of Hydrology*, 282(1-4), 56-75.

Hirpa, F. A., M. Gebremichael, and T. Hopson (2010), Evaluation of High-Resolution Satellite Precipitation Products over Very Complex Terrain in Ethiopia, *Journal of Applied Meteorology and Climatology*, 49(5), 1044-1051.

Hossain, F., and E. N. Anagnostou (2004), Assessment of current passive-microwave- and infrared-based satellite rainfall remote sensing for flood prediction, *Journal of Geophysical Research: Atmospheres*, 109(D7), n/a-n/a.

Huffman, G. J., D. T. Bolvin, E. J. Nelkin, D. B. Wolff, R. F. Adler, G. Gu, Y. Hong, K. P. Bowman, and E. F. Stocker (2007), The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales, *Journal of Hydrometeorology*, 8(1), 38-55.

Hwang, S., W. D. Graham, A. Adams, and J. Geurink (2013), Assessment of the utility of dynamically-downscaled regional reanalysis data to predict streamflow in west central Florida using an integrated hydrologic model, *Regional Environmental Change*, 13(S1), 69-80.

Jasper, K., J. Gurtz, and H. Lang (2002), Advanced flood forecasting in Alpine watersheds by coupling meteorological observations and forecasts with a distributed hydrological model, *Journal of Hydrology*, 267(1), 40-52.

Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. Xie (2004), CMORPH: A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution, *Journal of Hydrometeorology*, 5(3), 487-503.

Kanamaru, H., and M. Kanamitsu (2007), Fifty-Seven-Year California Reanalysis Downscaling at 10 km (CaRD10). Part II: Comparison with North American Regional Reanalysis, *Journal of Climate*, 20(22), 5572-5592.

Komma, J., C. Reszler, G. Blöschl, and T. Haiden (2007), Ensemble prediction of floods – catchment non-linearity and forecast probabilities, *Nat. Hazards Earth Syst. Sci.*, 7(4), 431-444.

Krinner, G., N. Viovy, N. de Noblet-Ducoudré, J. Ogée, J. Polcher, P. Friedlingstein, P. Ciais, S. Sitch, and I. C. Prentice (2005), A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, *Global Biogeochemical Cycles*, 19(1).

Kucera, P. A., E. E. Ebert, F. J. Turk, V. Levizzani, D. Kirschbaum, F. J. Tapiador, A. Loew, and M. Borsche (2013), Precipitation from Space: Advancing Earth System Science, *Bulletin of the American Meteorological Society*, 94(3), 365-375.

Kundzewicz, Z. W., L. J. Mata, N. W. Arnell, P. DÖLl, B. Jimenez, K. Miller, T. Oki, Z. ŞEn, and I. Shiklomanov (2008), The implications of projected climate change for freshwater resources and their management, *Hydrological Sciences Journal*, 53(1), 3-10.

Lakew, H. B., S. A. Moges, and D. H. Asfaw (2017), Hydrological Evaluation of Satellite and Reanalysis Precipitation Products in the Upper Blue Nile Basin: A Case Study of Gilgel Abbay, *Hydrology*, 4(3), 39.

Le Moigne, P. (2009), *SURFEX scientific documentation*, 211 pp.

Lettenmaier, D. P., D. Alsdorf, J. Dozier, G. J. Huffman, M. Pan, and E. F. Wood (2015), Inroads of remote sensing into hydrologic science during the WRR era, *Water Resources Research*, 51(9), 7309-7342.

Ly, S., C. Charles, and A. Degré (2011), Geostatistical interpolation of daily rainfall at catchment scale: the use of several variogram models in the Ourthe and Ambleve catchments, Belgium, *Hydrol. Earth Syst. Sci.*, 15(7), 2259-2274.

Mei, Y., E. N. Anagnostou, E. I. Nikolopoulos, and M. Borga (2014), Error Analysis of Satellite Precipitation Products in Mountainous Basins, *Journal of Hydrometeorology*, 15(5), 1778-1793.

Mei, Y., E. Nikolopoulos, E. Anagnostou, D. Zoccatelli, and M. Borga (2016), Error Analysis of Satellite Precipitation-Driven Modeling of Flood Events in Complex Alpine Terrain, *Remote Sensing*, 8(4), 293.

Nikolopoulos, E. I., E. N. Anagnostou, F. Hossain, M. Gebremichael, and M. Borga (2010), Understanding the Scale Relationships of Uncertainty Propagation of Satellite Rainfall through a Distributed Hydrologic Model, *Journal of Hydrometeorology*, 11(2), 520-532.

Overpeck, J. T., G. A. Meehl, S. Bony, and D. R. Easterling (2011), Climate Data Challenges in the 21st Century, *Science*, 331(6018), 700-702.

Peña-Arancibia, J., A. van Dijk, M. Stenson, and N. R. Viney (2011), *Opportunities to evaluate a landscape hydrological model (AWRA-L) using global data sets*.

Petrescu, A. M. R., L. P. H. van Beek, J. van Huissteden, C. Prigent, T. Sachs, C. A. R. Corradi, F. J. W. Parmentier, and A. J. Dolman (2010), Modeling regional to global CH₄ emissions of boreal and arctic wetlands, *Global Biogeochemical Cycles*, 24(4), n/a-n/a.

Richard G. Allen, Luis S. Pereira, Dirk Raes, and M. Smith (1998), Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56, *FAO Irrigation and drainage paper 56. Rome, Italy: Food and Agriculture Organization of the United Nations ISBN 92-5-104219-5*.

Sahlu, D., S. A. Moges, E. I. Nikolopoulos, E. N. Anagnostou, and D. Hailu (2017), Evaluation of High-Resolution Multisatellite and Reanalysis Rainfall Products over East Africa, *Advances in Meteorology*, 2017, 14.

Schellekens, J., et al. (2017), A global water resources ensemble of hydrological models: the earth2Observe Tier-1 dataset, *Earth Syst. Sci. Data*, 9(2), 389-413.

Schneider U., Ziese M., Becker A., Meyer-Christoffer A., and F. P (2015), Global Precipitation Analysis Products of the GPCC *Global Precipitation Climatology Centre (GPCC) Deutscher Wetterdienst, Offenbach a. M., Germany.*

Setegn, S. G., R. Srinivasan, A. M. Melesse, and B. Dargahi (2010), SWAT model application and prediction uncertainty analysis in the Lake Tana Basin, Ethiopia, *Hydrological Processes*, 24(3), 357-367.

Seyyedi, H., E. N. Anagnostou, E. Beighley, and J. McCollum (2015a), Hydrologic evaluation of satellite and reanalysis precipitation datasets over a mid-latitude basin, *Atmospheric Research*, 164-165(Supplement C), 37-48.

Seyyedi, H., E. N. Anagnostou, E. Beighley, and J. McCollum (2015b), Hydrologic evaluation of satellite and reanalysis precipitation datasets over a mid-latitude basin, *Atmospheric Research*, 164-165, 37-48.

Shiklomanov, I. A., and J. C. Rodda (2003), World water resources at the beginning of the 21st century, *Cambridge, UK: Cambridge University Press.*

Sutcliffe, J. V., and Y. P. Parks (1999), *The hydrology of the Nile*, xi + 179 pp. pp., IAHS Press, Wallingford.

Taylor, K. E. (2001), Summarizing multiple aspects of model performance in a single diagram, *Journal of Geophysical Research: Atmospheres*, 106(D7), 7183-7192.

Tong, X., D. Wang, V. P. Singh, J. C. Wu, X. Chen, and Y. F. Chen (2015), Impact of Data Length on the Uncertainty of Hydrological Copula Modeling, *Journal of Hydrologic Engineering*, 20(4), 05014019.

van den Hurk, B., and P. Viterbo (2003), The Torne-Kalix PILPS 2(e) experiment as a test bed for modifications to the ECMWF land surface scheme, *Global and Planetary Change*, 38(1), 165-173.

van Dijk, A., and L. Renzullo (2011), *The role of satellite observation in Australian water resources monitoring*, 10-15 pp.

Vörösmarty, C. J., and D. Sahagian (2000), Anthropogenic Disturbance of the Terrestrial Water Cycle, *BioScience*, 50(9), 753-765.

Vörösmarty, C. J., P. Green, J. Salisbury, and R. B. Lammers (2000), Global Water Resources: Vulnerability from Climate Change and Population Growth, *Science*, 289(5477), 284-288.

Wang, J., et al. (2011), The coupled routing and excess storage (CREST) distributed hydrological model, *Hydrological Sciences Journal*, 56(1), 84-98.

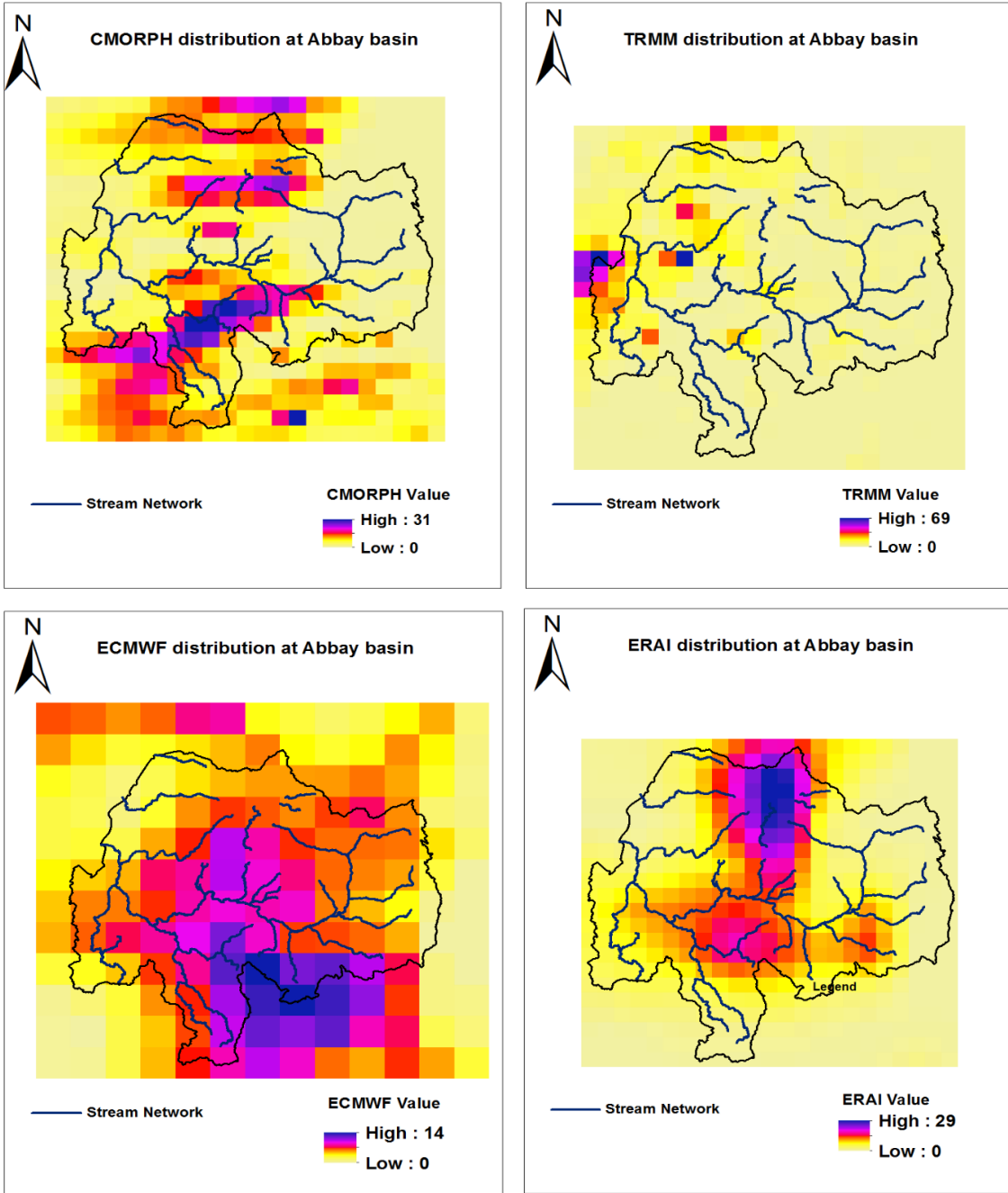
Wilk, M. B., and R. Gnanadesikan (1968), Probability plotting methods for the analysis for the analysis of data, *Biometrika*, 55(1), 1-17.

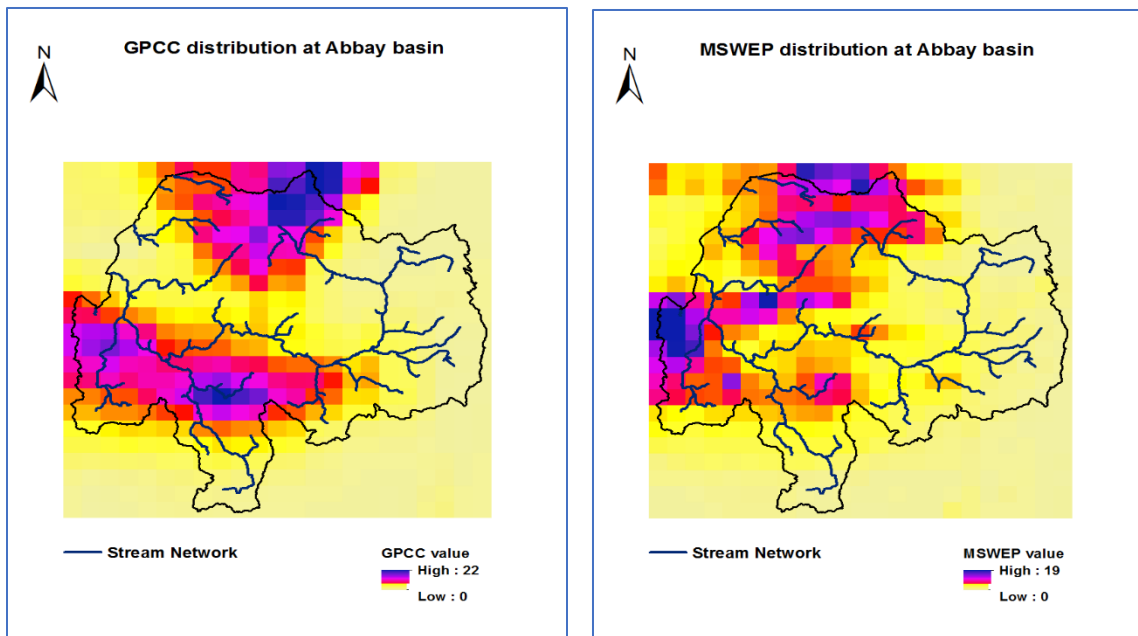
WWAP (2006), World Water Assessment Program, *The United Nations World Water Development Report 2*.

Xinyi Shen, and Y. Hong (2014), CREST Coupled Routing and Excess Storage CREST User Manual v2.1, *University of Oklahoma (OU) HyDROS Lab*(<http://hydro.ou.edu>).

Appendices

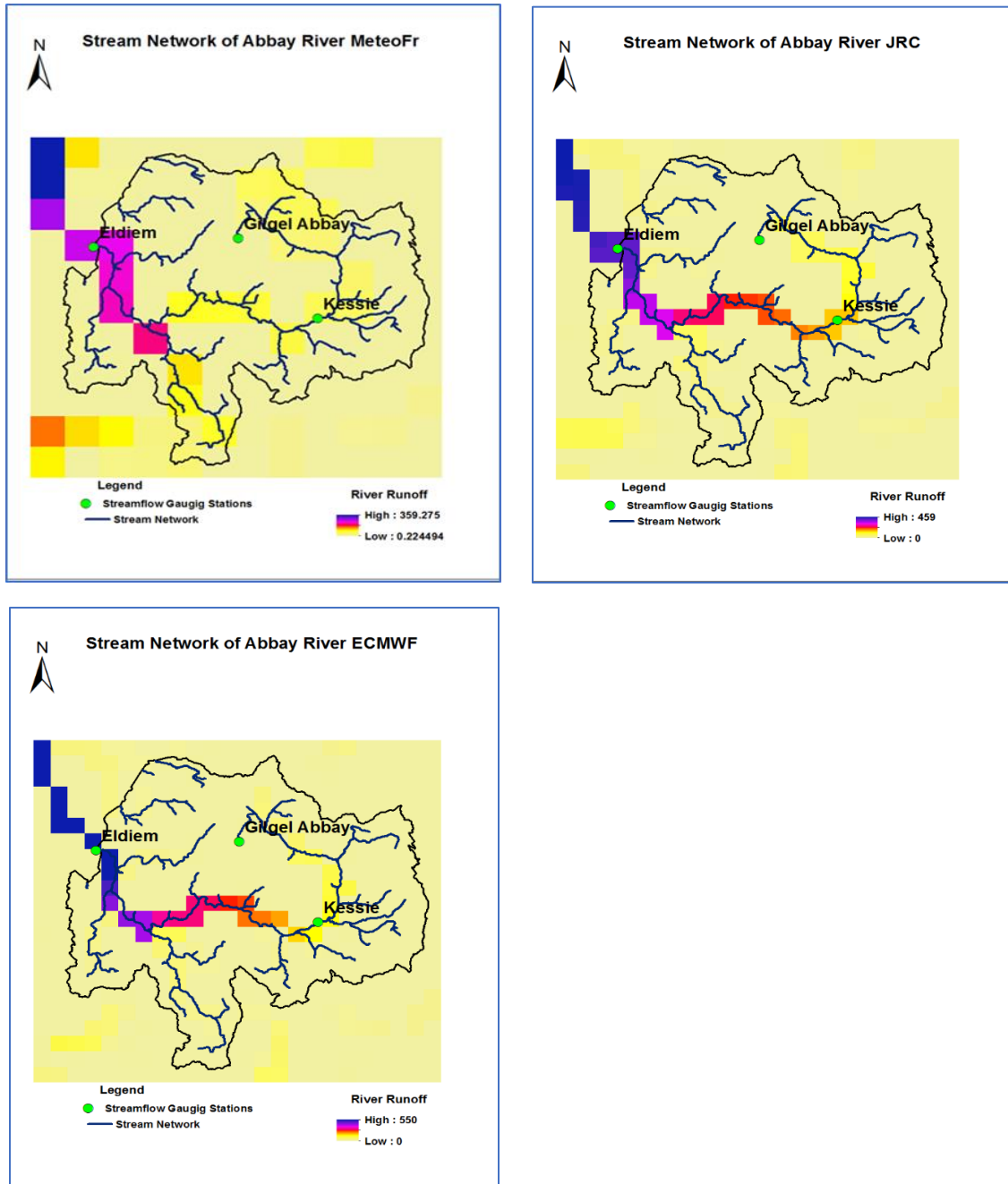
A. Global precipitation products distribution over the upper Blue Nile basin.





A 1 Spatial distribution of global precipitation products over the upper Blue Nile basin.

B. Water Resources Reanalysis (WRR2) river runoff products distribution over the upper Blue Nile basin.



B 1. Water Resources Reanalysis (WRR2, 0.25°) river runoff products with the natural stream network of Abbay basin

Publications

1. Lakew, H. B., S. A. Moges, and D. H. Asfaw (2017), Hydrological Evaluation of Satellite and Reanalysis Precipitation Products in the Upper Blue Nile Basin: A Case Study of Gilgel Abbay, *Hydrology* 2017, 4, 39; doi:10.3390/hydrology4030039.
2. Lakew, H. B., S. A. Moges, Anagnostou, E.N, Nikolopoulos, E.I and D. H. Asfaw, Evaluation of Global Water Resources Reanalysis Runoff Products for Local Water Resources Applications: A case of Upper Blue Nile Basin in Ethiopia. WARM-D-18-00167 (Under Review).