UCLA
UCLA Electronic Theses and Dissertations

Title

Evaluation of Short-Range and Medium- Range Precipitation Forecasts using Remote-Sensing Data

Permalink

https://escholarship.org/uc/item/9bg7q33w

Author

YUE, HAOWEN

Publication Date

2021

Peer reviewed|Thesis/dissertation

eScholarship.org
Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Los Angeles

Evaluation of Short-Range and Medium-Range Precipitation Forecasts using Remote-Sensing Data

A dissertation submitted in partial
satisfaction of the requirements for the
degree Doctor of Philosophy
in Civil Engineering

by

Haowen Yue

2021

© Copyright by

Haowen Yue

2021

ABSTRACT OF THE DISSERTATION

Evaluation of Short-Range and Medium-Range Precipitation Forecasts using Remote-Sensing Data

by

Haowen Yue
Doctor of Philosophy in Civil Engineering
University of California, Los Angeles, 2021
Professor Mekonnen Gebremichael, Chair

Accurate weather forecast information has the potential to improve water resources management, energy, and agriculture. This study evaluates the accuracy of recently developed short-range (0-18 hour) precipitation forecasts from the High-Resolution Rapid Refresh (HRRR) for selected extreme events over the US using NOAA's Gauge-corrected Multi-Radar/Multi-Sensor (MRMS-GC) radar-gauge merged rainfall observations. This study also evaluates the accuracy of medium-range (1-15 day) precipitation forecasts from the Global Forecast System (GFS) over transboundary river basins in Africa using NASA's Integrated Multi-satellitE Retrievals (IMERG) "Final Run" satellite-gauge merged rainfall observations. The assessment of HRRR on an hourly basis shows that there was a good agreement between the forecasted and observed precipitation in terms of temporal variability despite the forecasts tend to overestimate rainfall for hurricanes. Spatially, the forecasts were able to capture the general spatial pattern of hurricane driven events but failed to reproduce the characteristics of frontal storms. With regard to the effect of lead times, the 1-h lead forecasts have often lower accuracy than the other lead-time forecasts, while there was not much systematic difference in accuracy among the 2-h to 18-h lead-time forecasts. The evaluation of GFS reveals that the accuracy of forecasts varies a lot among different regions. GFS forecasts tend to overestimate precipitation in wet climatic regime but produce almost unbiased precipitation in dry regions. The GFS forecast accuracy decreases as the lead time increases, but the rate of decrement depends on the region. Aggregating the forecasts at temporal scales (1-day to 15-day) may increase or decrease the performance of GFS forecasts, depending on the region. We recommend exploring methods to increase the performance of short-range and medium-range forecasts, including post-processing techniques products before their application in water resources management.

The dissertation of Haowen Yue is approved.

David Yates

Timu Gallien

Steven Adam Margulis

William W. Yeh

Mekonnen Gebremichael, Committee Chair

University of California, Los Angeles

2021

Table of Content

List of Figures

List of Tables

ACKNOWLEDGEMENT

I wish to acknowledge the generous support and encouragement of my advisor Dr. Mekonnen Gebremichael in guiding me through my doctoral research, without whom this dissertation would not be possible. I thank Dr. David Yates for providing guidance and insightful suggestions. I also like to thank my dissertation committee member Dr. William W-G Yeh, Dr. Steven Margulis, and Dr. Timu Gallien for their valuable suggestions.

The dissertation contains the following published article:

Yue, H. and Gebremichael, M., 2020. Evaluation of high-resolution rapid refresh (HRRR) forecasts for extreme precipitation. *Environmental Research Communications*, *2*(6), p.065004. doi: 10.1088/2515-7620/ab9002.

Yue, H., Gebremichael, M. and Nourani, V., 2021. Evaluation of Global Forecast System (GFS) Medium-Range Precipitation Forecasts in the Nile River Basin. *Journal of Hydrometeorology*. accepted. doi: 10.1175/JHM-D-21-0110.1

Yue, H., Gebremichael, M. and Nourani, V., 2021. Performance of the Global Forecast System's Medium-Range Precipitation Forecasts in the Niger River Basin. *Hydrology and Earth System Sciences Discussions*, pp.1-31, accepted. doi: 10.5194/hess-2021-250

Author contribution is as follows:

H. Yue: data processing, data analysis, and manuscript preparation; M. Gebremichael: project oversight, method design, contribution to manuscript text; V. Nourani: method design, contribution to manuscript text.

The funding information is as follows:

Funding: This research was funded by NASA Precipitation Measurement Mission, grant number 80NSSC19K0688

VITA

EDUCATION

2016 B.Sc. (Water Project), Hohai University, China

2017 M.Sc. (Civil Engineering), University of California, Los Angeles

2017-2021 Graduate Student Researcher, Department of Civil and Environmental Engineering, University of California, Los Angeles

PUBLICATIONS

Yue, H. and Gebremichael, M., 2020. Evaluation of high-resolution rapid refresh (HRRR) forecasts for extreme precipitation. *Environmental Research Communications*, *2*(6), p.065004. doi: 10.1088/2515-7620/ab9002.

Yue, H., Gebremichael, M. and Nourani, V., 2021. Evaluation of Global Forecast System (GFS) Medium-Range Precipitation Forecasts in the Nile River Basin. *Journal of Hydrometeorology*. accepted. doi: 10.1175/JHM-D-21-0110.1

Yue, H., Gebremichael, M. and Nourani, V., 2021. Performance of the Global Forecast System's Medium-Range Precipitation Forecasts in the Niger River Basin. *Hydrology and Earth System* Sciences Discussions, pp.1-31, accepted. doi: 10.5194/hess-2021-250

CHAPTER 1

Introduction

1.1 Background and Motivation

Recent advances in forecasting have created opportunities to improve water resources management. The accuracy of weather forecasts has steadily improved over the years. The objective of the study is to identify the uncertainty in the most recent precipitation forecast models through evaluation with independent rainfall observations over continental United States and African river basins.

1.1.1 Precipitation Forecasts

Globally, precipitation forecasts from numerical weather prediction (NWP) models are becoming increasingly available to the public. Using real-time weather conditions as input to atmospheric models, NWP models predict the movement of weather systems to generate a variety of weather condition forecasts by representing the atmosphere as a dynamic fluid and solving for its behavior through the use of mechanics and thermodynamics (Cuo et al. 2011; Trenberth 1992; Buizza et al. 1999).

Depending on the lead time, precipitation forecasts can be grouped into three categories: short-range (0 - 120 hour), medium-range (1 - 15 day), and seasonal (1- 12 month) forecasts. While short-range forecasts are useful for flood early warning, medium-range precipitation forecasts and seasonal forecasts are useful for operational reservoir operation decisions, and water resources allocation and planning.

1.1.2 Application of Precipitation Forecasts

Precipitation forecasts with a variety of lead times have been widely used in water resources management worldwide, such as, flood/drought early warning systems (e.g., Cuo et al., 2011; Thielen et al., 2009), reservoir operation (e.g., Choong & El-Shafie 2015; Collischonn et al., 2007), and agriculture (e.g., Ingram et al., 2002; Schneider & Garbrecht 2003). Short-range and medium-range forecasts have been intensively used for flood warning, agriculture, and reservoir management (Patt et al. 2007; Breuer et al. 2010; Mase and Prokopy 2014; Pandya et al. 2015; Alexander et al. 2021), while seasonal forecasts are usually used for long-term reservoir management and drought prediction (Dutra et al. 2014; Mo & Lyon 2015; Yuan & Wood 2013).

In terms of flood warning, short-range and medium-range precipitation forecasts are usually integrated with hydrologic models to produce streamflow forecasts. For example, the National Water Model (NWM; NOAA 2015) ingests forcing from a variety of Numerical Weather Prediction forecast data including High Resolution Rapid Refresh (HRRR), Rapid Refresh (RAP), North American Mesoscale Nest (NAM-Nest), Global Forecast System (GFS) and Climate Forecast System (CFS) to provide streamflow forecasts for 2.7 million river reaches and other hydrologic information on 1km and 250m grids over the entire continental United States (CONUS). European Flood Alert System (EFAS; Thielen et al. 2009) incorporates output from ECMWF, the German Meteorological Service (DWD), and Consortium for Small-Scale Modeling Limited-Area Ensemble Prediction System (COSMO-LEPS) to increase preparedness for floods in European river basins. Other similar forecasting systems include the National Hydrometeorological Service (SCHAPI) from France, which merge short-range forecasts from radar and NWP models, and the operational medium-range hydrologic ensemble forecasting service, which uses ECMWF as inputs, from the Swedish Meteorological and Hydrological

Institute (SMHI).

Medium-range and seasonal forecasts are also widely applied in hydropower optimization and drought warning. For example, in a recent study, Koppa et al (2019) showed that the use of seasonal precipitation forecasts from NMME in reservoir planning of Omo Gibe dam in Ethiopia can increase annual hydropower generation by around 40%. Tang et al (2010) pointed out that using 10-day total precipitation forecasts from GFS would benefit reservoir optimization by increasing power generation by 114 GWh. Like the integrated flood warning system, seasonal forecasts from CFSv2, ECMWF, and NMME have been used as inputs to drive hydrologic models to predict future droughts (Fundel et al. 2013; Hao et al. 2018; Sheffield et al. 2014; Shukla et al. 2014). An example of such systems is the NCEP/EMC NLDAS Seasonal Hydrologic Forecast System developed by Princeton University and the University of Washington.

1.1.3 Need for Precipitation Forecasts Evaluation

Precipitation forecasting remains one of the most difficult challenges in NWP modeling because of enormous variability of the variables affecting the precipitation process (Golding 2000; Ebert et al. 2003). The performance of precipitation forecasts is affected by various factors such as forecast lead time, spatial and temporal scale, and climatic regimes (Cuo et al. 2011). The lack of understanding of precipitation forecast accuracy in different regions of the world blocks the path to utilizing forecast information in water resources management, energy and agriculture. In the Data and Methodology Section (Chapter 2), we will describe the current state of forecast evaluation.

1.2 Research Questions

The dissertation aims to address three research questions:

1. What is the accuracy of short-range precipitation forecasts for extreme precipitation in the United States?
2. What is the accuracy of medium-range precipitation forecasts over African river basins?

1.3 Dissertation Overview

The dissertation is divided into twelve chapters. Chapter 1 provides an introduction. Chapter 2 provides data and methodology. Chapter 3 presents evaluation of short-range forecasts derived from the High-Resolution Rapid Refresh (HRRR) forecasts over US. Chapter 3 through Chapter 11 present the evaluation of medium-range forecasts derived from Global Forecast System (GFS) over African River Basins – each chapter focuses on one river basin. Chapter 12 provides conclusions.

CHAPTER 2

Data and Method

2.1 Data and Methodology for Evaluation of Short-Range Forecasts in the U.S.

2.1.1 HRRR Forecast System

The High-Resolution Rapid Refresh (HRRR) model of the National Weather Service (NWS) provides state-of-the-art, short-range (lead times ranging from 1 h to 18 h), high-resolution (3 km, hourly) QPFs across the continental United States (CONUS). With the growing need for frequently-updated short-range weather guidance, the fine-resolution HRRR was developed and introduced into the operational model suite at the NOAA/National Centers for Environmental Prediction (NCEP) in September 2014. The HRRR has been undergoing continuous model improvement, almost on a yearly basis. The HRRR model is a real-time, hourly-updated, convection-allowing, storm resolving model running at 3 km horizontal resolution with 50 vertical levels over the CONUS. Its domain is nested within the 13 km Rapid Refresh mesoscale model, which also provides boundary conditions (Benjamin et al 2016). The HRRR is built upon the Advanced Research version of Weather Research Forecast (WRF-ARW) model with the following physics options: the Goddard shortwave radiation scheme (Chou and Suarez 1994), the Rapid Radiative Transfer Model longwave radiation scheme (Mlawer et al 1997), the Rapid Update Cycle smirnova land surface model (Smirnova et al 1997), the Mellor-Yamada-Nakanishi-Niino boundary layer parameterization (Nakanishi and Niino 2004), and the Thompson mixed-phase microphysics scheme (Thompson et al 2008). Initial fields are created using 3D-VAR data assimilation. The latent heating profile are calculated as a function of radar reflectivity (which came from the NWS WSR-88D network that is used to create the Stage IV products) which is assimilated at 3 km resolution every 15 min (Benjamin et al 2016). Detailed information on model history and physics are provided in Benjamin et al (2016).

The HRRR model provides operational forecasts at 3-km and hourly resolution, with lead times ranging from 1 h to 18 h. It also provides experimental sub-hourly products in order to test updated model performance and measure performances resulting from updates. Most recent versions are HRRRv2 available since July 2016, and HRRRv3 forecasts available since 2018. These data are available to the public via the HRRR Archive managed by the University of Utah (Balylock et al., 2017). The NWS' operational hydrologic modeling platform NWM first downscales the HRRR forecasts from 3 km to 1 km by bilinear interpolation methods, and uses the downscaled forecast as forcing input to generate short-range hydrological forecasts at lead times ranging from 1 h to 18 h (NOAA, 2016). Studies evaluating the accuracy of QPF, particularly during extreme events, are very limited as shown below.

Gowan et al (2018) assessed the performance of forecasts by HRRR and coarse-resolution NWP models (GFS, NAM-3 km CONUS nest and NCAR Ensemble) over mountainous western US by comparing the forecasts to precipitation observation from the and precipitation analyses from the Parameter-Elevation Regressions on Independent Slopes Model (PRISM), and reported that the HRRR forecasts outperformed the coarse-resolution NWP models partly due to better representation of topography in the HRRR model. Pinto et al (2015) reported that the HRRR predicted too many mesoscale convective systems (MCSs) over the Great Plains and too few MCSs over the southeastern United States, and that the skill of the HRRR at predicting specific MCS events increased between 2012 and 2013, coinciding with changes in model physics and data assimilation technique. Cai and Dumais (2015) evaluated three weeks of HRRR forecasts during the summer of 2010 over the eastern United States, and reported that the HRRR model was able to capture the overall convective storm characteristics but with significant bias which varies substantially by region.

Bytheway and Kummerow (2015) estimated the performance of the 2013 experimental version of the HRRR forecasts for warm-season convective storms over central US by comparing the forecasts to Stage IV NEXRAD Radar rainfall products, and reported the following results: a southward displacement of the storm structure by about 30 km, large overestimation in intensity and areal extent of precipitation, overestimation of heavy rain, and relatively better forecast skill for the 3-h lead forecast. Two years later, Bytheway et al (2017) performed a similar study using the research versions of latter-year (2013–2015) HRRR forecasts, and concluded overall improvements in the bias for maximum and mean rainfall intensity and storm location as a result of model upgrades, compared to the 2013 forecasts.

Ikeda et al (2013) investigated the accuracy of winter-season HRRR forecasts with a particular focus on areal extent of precipitation and the timing and location of regions of different precipitation phases during the 2010/11 cold season by comparing the forecasts to observations from the Automated Surface Observing System (ASOS) station network across the Eastern United States and revealed that larger synoptically forced weather systems were better predicted than smaller weather systems. Ikeda et al (2017) examined HRRR model's ability to forecast the surface precipitation phase for cold-storm cases from the 2013/14 and 2014/15 winter seasons over the east of the Rocky Mountains and concluded that the HRRR model was able to represent the overall vertical thermodynamic structure in the mixed-phase precipitation regions. Seo et al (2018) evaluated the performance of HRRR forecasts for the September 20–23, 2016, frontal storms over Eastern Iowa by comparing the forecasts to the Multi-Radar/Multi-Sensor Gauge-Corrected (MRMS-GC) products and reported different behavior for different statistics: while the forecast skill decreased with lead time (maximum skill achieved at 1-h lead), the bias and stand error improved with lead time.

For our analyses, we have used the recent HRRR forecasts, namely, the HRRRv2 forecasts for the frontal storms and hurricane Harvey and the HRRRv3 forecasts for the hurricane Florence. In order to provide results relevant for NWM, we have downscaled the hourly HRRR forecasts from 3 km to 1 km implementing identical downscaling methods used in NWM.

2.1.2 The Reference Observation Data MRMS-GC

The gauge-corrected Multi-Radar/Multi-Sensor (MRMS-GC) quantitative precipitation estimates (QPE) are produced operationally at the National Centers for Environmental Prediction (NCEP) and distributed to NWS forecast offices and several external agencies. The MRMS-GC ingests radar reflectivity data from the WSR-88D network resulting in a spatial domain covering the CONUS. Radar-based estimates of precipitation are adjusted using a network of approximately 7000 gauges from the Hydrometeorological Automated Data System (HADS) network (Kim et al 2009). Data from the NWP model Rapid Update Cycle (RUP) are used in quality controlling the rain gauge data. The resulting MRMS-GC QPEs are hourly rainfall rates with a spatial resolution of 0.01° resolution. Detailed information on MRMS-GC can be found in REFs. The MRMS-GC product has a history of use as the reference product for validation of models and satellite products (e.g., Gourley et al 2017, Gebregiorgis et al 2017, Smalley et al 2017, Seo et al 2018). In this study, we used the MRMS-GC QPEs as reference to validate the HRRRv2 and HRRRv3 forecasts.

2.1.3 Evaluation Methods

The evaluations of HRRR forecasts were carried out for various quality features. These include: (1) hourly comparisons of domain-averaged precipitation, fraction of domain covered by rain, and conditional coefficient of variation (i.e. standard deviation of rainfall within the domain normalized by domain-averaged precipitation – only taking into account pixels where rainfall occurred), (2) spatial maps of accumulated rainfall over the entire storm duration, and (3) The geometric features of accumulated rainfall using the Amplitude Structure Location (SAL) method.

The SAL method (Wernli et al., 2008) evaluates storm features (size, variability, and location) by identifying precipitation objects in both the forecast and the observed storm at a given time, and decomposing differences (i.e. errors) into three components. The errors are normalized by the size of the domain and domain-wide accumulation such that results from different domain sizes and rainfall accumulation can be compared. The amplitude (A) component (between –2 and +2) corresponds to the normalized difference of the domain-averaged precipitation values. $A=0$ denotes perfect forecast, $A>0$ indicates overestimation, $A<0$ indicates underestimation of domain-averaged precipitation. The values $A=0.4$, 0.67, 1 indicate overestimation of domain-averaged precipitation by a factor of 1.5, 2, and 3, respectively. Along the same lines, $A = -1$ indicates underestimation by a factor of 3. The Structure (S) component (between –2 and +2) captures information about the size and shape of precipitation objects. The value $S = 0$ indicates perfect field, $S > 0$ indicates widespread precipitation forecast in a situation of localized events, while $S < 0$ indicates too small precipitation objects or too peaked objects, or a combination of these factors. The location component L (between 0 and 2) consists of two parts. One part measures location differences in centers of mass for the domain-wide observed and forecast fields; the other part accounts for location differences of all objects weighted by their integrated precipitation. L = 0 indicates a forecast field, where both the center of mass as well as the averaged distance between the precipitation objects and the center of mass agree with the observations. The ASL equations are provided in Wernli et al. (2008).

2.2 Data and Methodology for Evaluation of Medium-Range Forecasts in Africa

2.2.1 Global Forecast System (GFS)

The Global Forecast System (GFS) is a global numerical weather prediction system run by the U.S. National Weather Service (NWS). The GFS forecast products with a resolution of 0.25° by 0.25° are obtained from National Center for Atmospheric Research (NCAR) Research Data Archive (RDA) GFS Historical Archive (NCEP 2015). The GFS is run four times a day at 00, 06, 12, and 18 UTC. One of the GFS model output variables is accumulated precipitation, where the precipitation forecasts are accumulations starting from the model run time. We obtained the 1-day lead daily rainfall forecast by subtracting the 24-hour rainfall accumulation forecast from the 48-hour rainfall accumulation forecast. Similarly, in order to obtain the 5-day lead daily rainfall forecast, we subtracted the 120-hour rainfall accumulation forecast from the 144-hour rainfall forecast. We only considered the model runs at 00 UTC.

The GFS model went through a major upgrade, and its version-15 forecasts are available since June 12, 2019. In version 15, the Finite Volume Cubed Sphere dynamical model (FV3) replaced the Global Spectral Model (GSM) as the core model. In the GSM model, the horizontal resolutions were T1543 (12.5km) from 0 to 240 hours (0-10 days) and T574 (~34km) from 240 to 384 hours (10-16 days) (NCEP 2021a). However, in the FV3 model, the horizontal resolution of the model is about 13 km for days 0-16 (NCEP 2021b). The model runs are re-gridded to produce precipitation forecasts at 0.25° resolution (NCEP 2015).

Developed by the Geophysical Fluid Dynamics Laboratory (GFDL) (Putman and Lin, 2007), the key components of FV3 include: (1) the Rapid Radiative Transfer Method for GCMs (RRTMG) scheme for shortwave/longwave radiation (Mlawer et al. 1997; Iacono et al. 2000; Clough et al. 2005), (2) the Hybrid eddy-diffusivity mass-flus (EDMF) scheme for Planetary Boundary Layer (PBL) (NCEP, 2019a), (3) the Noah Land Surface Model (LSM) scheme for land surface option (Chen et al. 1997), (4) the Simplified Arakawa-Schubert (SAS) deep convection for cumulus parameterization (Arakawa et al. 1974; Grell 1993), and (5) an advanced GFDL microphysics scheme for microphysics (NCEP, 2019b).

However, GFS forecasts, especially version 15 is not well evaluated over Africa. Using observations collected over the eastern Pacific during the El Niño Rapid Response (ENRR) field campaign, Wang et al. (2019) reported that GFS forecasts have difficulty to capture the location and magnitude of heavy rain rates. Sridevi et al. (2018) evaluated the performance of GFS in India by using rain gauge and satellite rainfall product and reported that the GFS forecast shows some skills in 1-day and 2-day lead times, but low skills from 3-day onwards. Lien et al. (2016) compared the global statistical properties of GFS forecasts and Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) (Huffman et al. 2007) observations and reported that the GFS model has positive bias in precipitation amount compared to TMPA observations, and that the GFS forecasts have large random errors at higher resolutions, especially for convective precipitation. Jiang et al. (2015) evaluated the accuracy of GFS precipitation forecasts in China, U.S., and Australia using rain gauge data, and reported overestimation of light rain and underestimation of moderate and heavy rain, which they attributed to the errors arising from not accounting for the Aero-sol-Cloud Interactions in the GFS model.

2.2.2 IMERG

Integrated Multi-satellitE Retrievals for GPM (Global Precipitation Measurement) (IMERG) precipitation products are available at 30-min temporal resolution and 0.1° spatial resolution. IMERG uses data from the GPM satellite constellation, including a Ku/Ka-band Dual-frequency Precipitation Radar (DPR), a multi-channel GPM Microwave Imager (GMI), multiple partner passive microwave (PMW) instruments, and thermal infrared (IR) information from geostationary satellites (Huffman et al. 2019a). Details of IMERG algorithm are available at Huffman et al (2019b).

Currently, IMERG produces three types of products: “Early Run” or “Early” product, with a lag time of about 4 hours; “Late Run” product, with a lag time of about 14 hours, and “Final Run” or “Final” product, with a lag time of about 3.5 months (Huffman et al., 2015). The main difference in algorithm between IMERG Early and IMERG Final lies in the use of rain gauge information (GPCC monthly gauge-analysis) in IMERG Final for bias correction. The latest version (V6B) of IMERG datasets have been accessed from the NASA’s Earth Data Goddard Earth Sciences Data and Information Services Center (GES DISC) web portal. In this study, we use IMERG Final as a reference to evaluate both GFS forecasts and IMERG Early and other satellite products.

A number of studies have been conducted to validate IMERG product over Africa. For example, Sahlu et al. (2016) evaluated the accuracy of IMERG Final in comparison to data from 37 experimental rain gauge stations established by Gebremichael et al (2014) in the Blue Nile region, and reported bias under 5%. Dezfuli et al. (2017a) evaluated the performance of IMERG Final against two, high-resolution, experimental rain gauge station data provided by the Trans-African Hydro-Meteorological Observatory network (TAHMO; van de Giesen et al. 2014), and reported the capability of IMERG Final to represent well the diurnal cycle of rainfall. Validated against TAHMO network, Dezfuli et al. (2017b) showed that IMERG Final is able to capture the propagation of large Mesoscale Convective Systems, a significant advantage over its predecessor's satellite rainfall product, known as the Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA). Maranan et al. (2020) compared IMERG Final products against experimental rain gauge station data in the moist forest region of Ghana, West Africa, and showed that IMERG Final datasets are able to capture monthly rainfall with a very high correlation coefficient.

2.2.3 CHIRPS

The Climate Hazard's group Infrared Precipitation with Stations (CHIRPS) dataset, a merged satellite-gauge product, is available at a spatial resolution of 0.05° and a temporal resolution of 1-day, with a data latency period of about 3 weeks. CHIRPS rainfall products are obtained from the CHIRPS webpage. The CHIRPS products are derived primarily from thermal infrared data using the cold cloud duration (CCD) approach, calibrated using TRMM Multi-satellite Precipitation analysis (TMPA 3B42 v7; Huffman et al. 2007) precipitation datasets by local regression, and include rain gauge station data from multiple sources (regional and national meteorological services). Details of CHIRPS algorithm are available at Funk et al. (2015).

Fenta et al. (2018) evaluated the accuracy of CHIRPS by comparing them to data from 11 rain gauge stations in the Lake Tana (i.e. head watershed of the Blue Nile) watershed, and reported bias under 10%. Similar performance results were obtained for CHIRPS by other researchers using different rain gauge networks in different parts of the Blue Nile basin (Abdelmoneim et al. 2020; Belete et al. 2019). Satgé et al. (2020) evaluated the accuracy of a number of gridded precipitation datasets over West Africa through comparison against rain gauge station data, and reported that CHIRPS provides reliable estimates at both daily and monthly timescales, while other satellite rainfall products considered (CMORPH, PERSIANN, GSMaP, ARC, and TAMSAT) and all atmospheric reanalysis products considered (MERRA and JRA) are deemed unreliable. Furthermore, they reported that the satellite products that incorporated rain gauge information outperform satellite-only products.

2.2.4 Evaluation Methods

IMERG Final rainfall products are used in this study as reference to evaluate the performance of GFS precipitation forecasts. The comparison period is 15 June 2019 to 15 June 2020 to match the period for which the version-15 of GFS model forecasts is available. The spatial resolutions of the forecast and satellite products are different: 0.25°(GFS), 0.10° (IMERG Final and IMERG Early), and 0.05° (CHIRPS). The temporal resolutions of the satellite products are: 30-minute (IMERG Final and IMERG Early) and daily (CHIRPS). Our comparison is mostly based on sub-basin (i.e. watershed for each dam) average values, in which case we average all the datasets to the sub-basin spatial scale. In some cases, where we compare the spatial patterns of rainfall, we resample both IMERG products and CHIRPS to 0.25° using the bilinear interpolation technique to match the spatial resolution of GFS.

For evaluation metrics, we used the modified Kling-Gupta Efficiency (KGE; Gupta et al. 2009; Kling et al. 2012) and its components: Bias Ratio (BR), correlation (R), and variability ratio (γ). KGE measures the goodness-of-fit between estimates of precipitation forecasts and reference observations as:

$$KGE = 1 - \sqrt{(R-1)^2 + (BR-1)^2 + (\gamma-1)^2},$$

$$BR = \frac{\mu_f}{\mu_o},$$

$$\gamma = \frac{CV_f}{CV_o},$$

where R is the linear correlation coefficient between forecasted and observed precipitation, BR is the bias ratio, γ is the variability ratio, μ is the mean precipitation, CV is the coefficient of variation, and the indices f and o represent forecasted and observed precipitation values, respectively. KGE values range from -∞ to 1, with values closer to 1 indicating better model performance. Towner et al. (2019) suggested the following classifications: "Good" ($KGE \geq 0.75$), "Intermediate" ($0.75 \geq KGE \geq 0.5$), "Poor" ($0.5 \geq KGE > 0$), and "Very poor" ($KGE \leq 0$). The BR values greater than 1 indicate a positive bias whereby forecasts overestimate precipitation relative to the observed data, while values less than 1 represent an underestimation. The γ values greater than 1 indicate that the variability in the forecast time series is higher than that observed, and values less than 1 show the opposite effect. The R measures the strength and direction of the linear relationship between the forecast and observed values, and to what extent the temporal dynamics of observed rainfall is captured in the forecasts. The correlation values of 0.6 or more are considered to be skillful (e.g., Alfieri et al. 2013). In addition, the root mean-square-error normalized by reference precipitation mean (NRMSE) was also used.

CHAPTER 3

Evaluation of Short-Range Forecasts (HRRR) over CONUS

This chapter has been published in the Environmental Research Communications. © 2020 The Author(s). Published by IOP Publishing Ltd. Used with permission.

Yue, H. and Gebremichael, M., 2020. Evaluation of high-resolution rapid refresh (HRRR) forecasts for extreme precipitation. *Environmental Research Communications*, *2*(6), p.065004. doi: 10.1088/2515-7620/ab9002.

3.1 Storm Cases

Given our focus on extreme storms, we selected five disastrous storms from recent years as reported by the USGS Flood reports. These are: (1) the September 21–23, 2016, frontal storms in Iowa, (2) the April 28-May 1, 2017, frontal storms in the Southern Midwestern US, (3) the August 25–31, 2017, Hurricane Harvey storms in Texas, (4) the September 13–17, 2018, and (5) the September 4–6, Hurricane Dorian storms in the Carolinas., Hurricane Florence storms in the Carolinas. These cases allowed us to examine the uncertainties in forecasts during frontal storms as well as hurricanes in different parts of the U.S. Table 3.1 presents some of the key characteristics of these storms. The frontal system in Iowa had an accumulation of 91 mm over 69 h for the Cedar watershed (watershed area of 49,140 sq. km), and the frontal system in southern Midwestern had 158 mm of rainfall over 95 hours for the Meramec watershed (18,630 sq. km). The hurricanes had higher rainfall and lasted longer: hurricane Harvey in Texas resulted in 501 mm over 157 hours for the Lower Sabine Watershed (17,856 sq. km), Hurricane Florence had 496 mm over 128 hours for Lumber watershed (9,790 sq. km), and Hurricane Dorian had 109 mm over 62 hours for Caper Fear watershed (9,120 sq. km).

Table 3.1. Some of the key characteristics of the selected storms over nested watersheds

Watershed	Domain Size (km2)	Elevation (m)	Major Land Cover	Total Rainfall (mm)	Peak hourly rainfall (mm)	Duration (Hour)
The September 21-23, 2016, frontal storms in Iowa						
Cedar	49140	326	Cultivated Crops	91	5.97	69
The April 28-May 1, 2017, frontal storms in the Southern Midwestern US						
Flint	819	350	Deciduous Forest	159	25.83	56
Big	7650	303	Deciduous Forest	159	9.90	91
Gasconade	9108	286	Deciduous Forest	140	9.85	77
Eleven Point	9486	260	Deciduous Forest	109	10.81	68
Illinois	11025	355	Deciduous Forest	126	14.26	84
Meramec	18630	278	Deciduous Forest	158	8.94	95
The August 25-31, 2017, Hurricane Harvey storms in Texas						
Cypress	4104	73	Pasture/Hay	608	19.12	121
WF San Jacinto	9180	80	Pasture/Hay	554	17.86	139
Navidad	10260	86	Pasture/Hay	346	11.30	112
Lower Sabine	17856	86	Shrub/Scrub	501	12.93	175
The September 13-17, 2018, Hurricane Florence storms in the Carolinas						
Rockfish	1134	92	Cultivated Crops	334	18.24	109
Black	1450	80	Cultivated Crops	222	11.58	76
Neuse	5304	29	Cultivated Crops	316	12.91	107
Cape Fear	9120	28	Cultivated Crops	496	12.76	102
Lumber	9790	118	Cultivated Crops	333	11.78	128
The September 4-6, 2019, Hurricane Dorian storms in the Carolinas						
Rockfish	1134	92	Cultivated Crops	58	7.14	27
Black	1450	80	Cultivated Crops	49	5.47	29
Neuse	5304	29	Cultivated Crops	119	15.79	34
Cape Fear	9120	28	Cultivated Crops	109	6.82	62
Lumber	9790	118	Cultivated Crops	66	5.44	42

3.2 Results and Discussion

3.2.1 The September 21-23, 2016, Frontal Storms in Iowa

A tropical air mass interacting with a stationary front triggered several rounds of heavy storms in northern Iowa and southern Minnesota during September 21–23, 2016 (Seo et al 2018). Figure 2.1a shows spatial map of accumulated rainfall accumulation during the entire storm period over a selected domain. Most of the domain was covered by rain. There was a supercell, where rainfall ranged from 250 mm to 300 mm at the core of the cell and from 200 mm to 250 mm at the peripheries. Let us first examine how the spatial pattern of the storm was portrayed by the forecasts. The spatial maps of accumulated rainfall for HRRR forecasts are shown in figures 1(b)–(e). The forecasts accurately showed rainfall occurrence for most of the domain. However, there were differences in the spatial pattern of rainfall: the 1-h lead forecast captured the observed supercell but overestimated its spatial coverage, the 6-h and 18-h lead forecasts missed the supercell, and the 12-h lead forecast displaced the supercell to the northeast and also increased its size.

In addition to spatial maps, hourly time series of precipitation patterns were evaluated. For this purpose, the domain-averaged precipitation values were evaluated at each hour during the storm period in figure 2(a). According to the observations, the storm lasted 72 h, and had a wave pattern with three peaks. The first and smaller peak was at hour 7 (counting from the beginning of the storm) with a peak rainfall of 1.12 mm h^{-1}, the second and largest peak at hour 34 with a peak rainfall of 5.78 mm h^{-1}, and the third and moderate peak occurred at hour 58 with a peak rainfall of 2.74 mm h^{-1}. Overall, the forecasts captured well the wave pattern with three peaks. However, the forecasts varied in how they forecasted the hydrographs for the peak events.

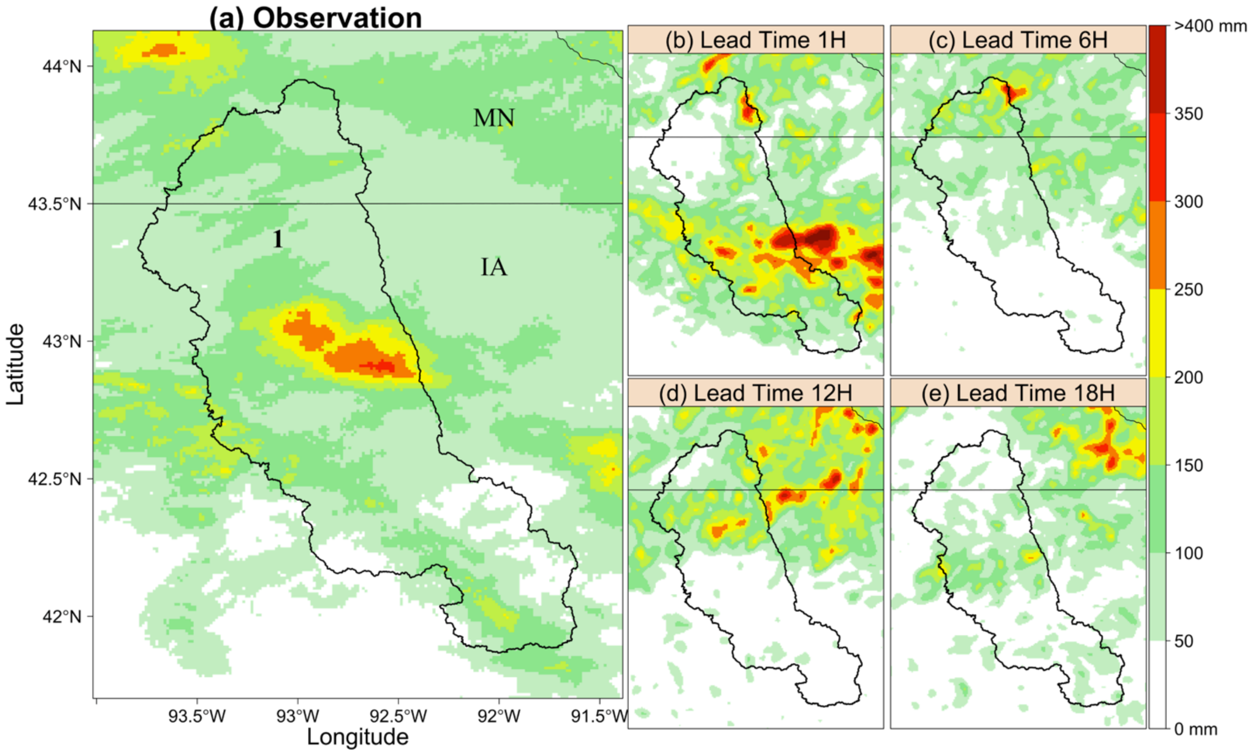


Figure 3.1. Precipitation accumulation map (3-days) of frontal system driven storms in Iowa: (1) Cedar River basin; (a) Observation (b) Forecasts at Lead Time 1H; (c) Forecasts at Lead Time 6H; (d) Forecasts at Lead Time 12H; (e) Forecasts at Lead Time 18H.

The second and largest peak (recall peak rainfall of 5.78 mm h−1 for observed rainfall) was well-forecasted with only errors of +9% (1-h lead forecast), −15% (6-h), +13% (12-h), +15% (18-h). However, there were differences among the forecasts in the rising and falling limbs of the hyetograph for this event: the 1-h lead forecast displaced the storm by about 4 h (rising/falling limbs and peak were 4 h late than the observations); the 6-h lead forecast correctly reproduced the rising and falling limbs; the 12-h lead forecast showed unusual spike at the beginning of the storm, but captured well the remaining portion of the hyetograph.; the 18-h lead forecast missed almost the rising limb of the hyetograph, but captured the recession limb. For the third and moderate peak (recall peak rainfall of 2.74 mm h−1 for observed rainfall), the 1-h lead forecast overestimated the peak (by about 39%), while the remaining forecasts substantially underestimated the peak (by 70% for 6-h, by 42% for 12-h, by 63% for 18-h lead forecasts). For this event, the 1-h lead forecast captured well the hydrograph with no delays, while the 6-h lead forecast missed the hydrograph. For the first and smallest-peak event (recall peak rainfall of 1.12 mm h^{-1} for observed rainfall), the 1-h forecast best matched the observation albeit with some underestimation, while the remaining forecasts resulted in large underestimation of the peak.

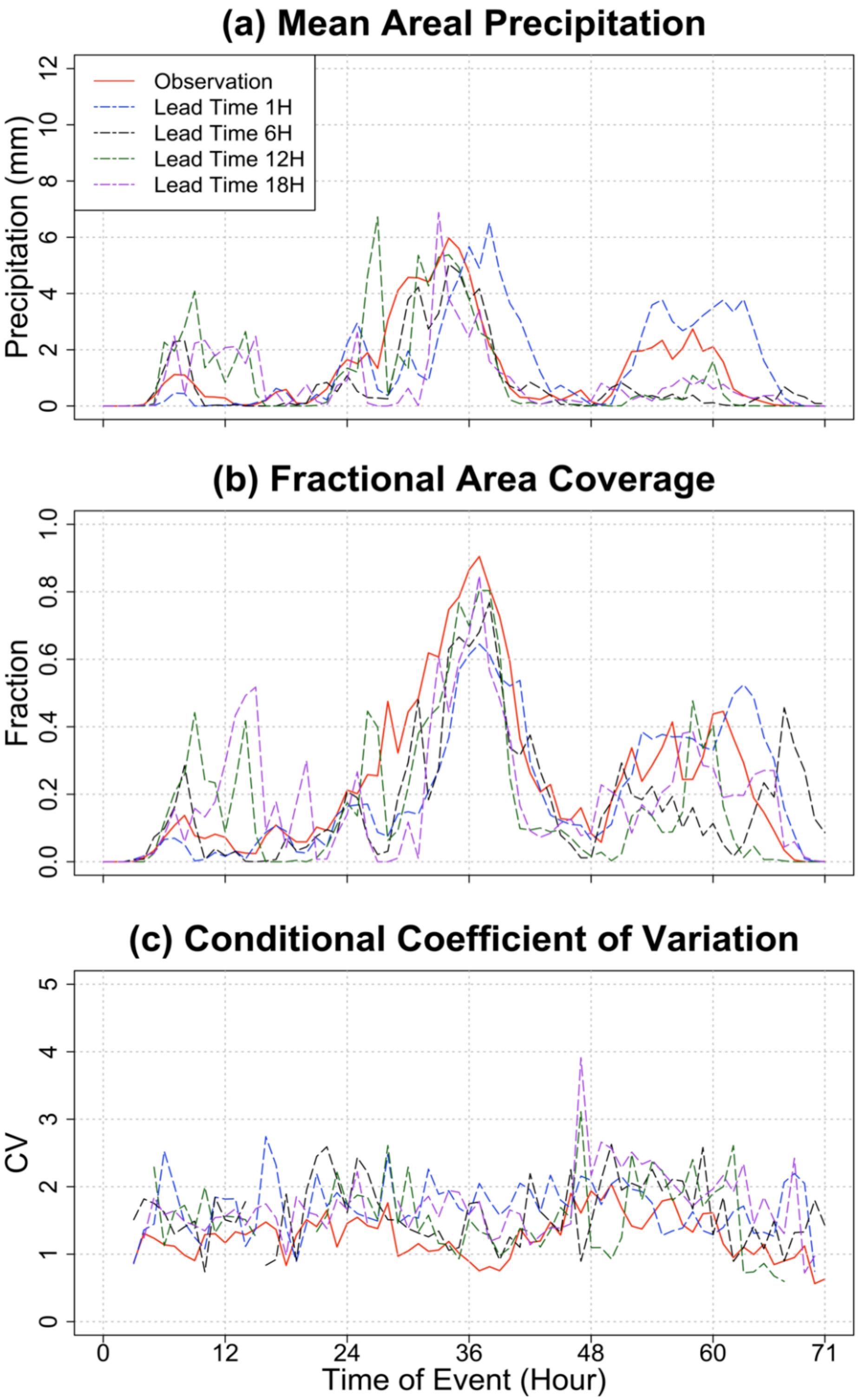


Figure 3.2. Event overview of frontal system driven storms in Iowa: (a) (a) Time series of mean areal precipitation; (b) Time series of fractional area coverage; (c) Time series of conditional coefficient of variation.

Next, we evaluated the fraction of the domain covered by rain (f_{rain}) by dividing the number of pixels (1-km grids) with rainfall intensity larger than zero to the total number of pixels within the domain. Figure 3.2b shows the time series of f_{rain} for the observation and forecasts. According to the observations, the temporal pattern of f_{rain} follows the temporal pattern of the domain-averaged rainfall discussed above, indicating that the bigger the storms the larger the area they cover. The forecasts reproduced this temporal pattern of f_{rain}. However, there were some differences in the actual magnitudes of f_{rain}: for the second and largest-peak event, all the forecasts underestimated the areal coverage; for the third and moderate-peak event, the 1-h lead forecast agreed well with the observations, while the remaining lead-time forecasts underestimated the spatial coverage; and for the first and smallest-peak event, the 1-h lead forecast did well, while the remaining lead forecasts overestimates the areal coverage. Overall, the performance of the lead-time forecasts in forecasting the areal coverage is similar to their performance in forecasting the domain-averaged precipitation.

Finally, we evaluated the conditional coefficient of variation of rainfall (CV_{con} defined as standard deviation of rainfall within the domain normalized by area-averaged rainfall, only considering pixels with rainfall exceeding zero). According to the observations, CV_{con} varied mostly between 0.9 and 2 and its temporal pattern was not associated with the magnitude of the storm. All the forecasts reproduced the temporal pattern of CV_{con} observed in the observed rainfall fields, however, they showed higher values at almost all hours, indicating that the forecasted fields had higher spatial variability than the observed rainfall fields.

3.2.2 The April 28-May 1, 2017, Frontal Storms in the Southern Midwestern US

A convergence of two fronts (a warm front that extended from southeast Missouri across west-central Arkansas into south-central Oklahoma moving north, and a tropical moisture from the Gulf of Mexico moving north) produced thunderstorms that yielded abundant precipitation over the Southern Midwestern US during the period April 28th - May 1st, 2017 (Heimann et al 2018). In figure 3, spatial maps of accumulated rainfall forecasts were compared to those of observations to get a first impression of the quality of the forecasts. According to the observations (figure 3(a)), there was a wide strip of moderate-intensity rain in the southwest-northeast direction, with large rainfall accumulation around the center of the domain. The 1-h lead forecast (figure 3(b)) placed the storm to the northwest, and produced quite a large number of scattered, localized, high-intensity rainfall events than what the observations show. The 6-h lead-time forecast (figure 3(c)) produced more-scattered, high- intensity, rainfall fields in the northeast of the domain, while significantly underestimating the large rainfall field at the center of the field, and placing the storms in the southwest slightly to the north. The 12-hour lead forecast (figure 3(d)) tended to underestimate observed precipitation across the domain. The 18-h lead forecast (figure 3(e)) captured the large rainfall observed around the center of the domain but showed some scattered high-intensity rainfall fields in the southern part of the domain which did not appear in the observations.

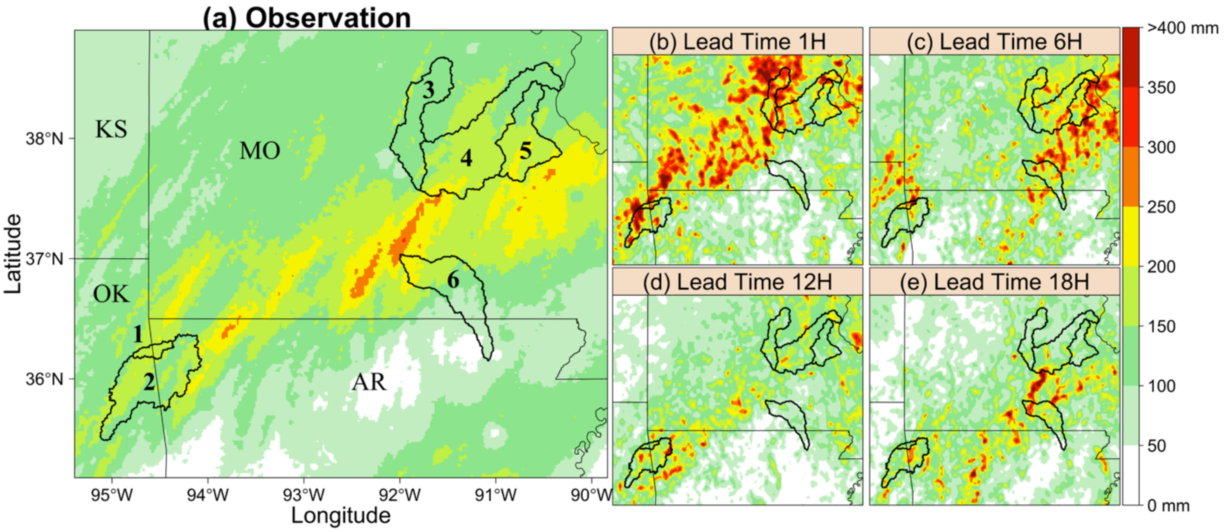


Figure 3.3. Precipitation accumulation map (3-days) of frontal system driven storms in Southern Midwestern US: (1) Flint creek basin; (2) Illinois River basin; (3) Gasconade River basin; (4) Meramec River basin; (5) Big River basin: (6) Eleven Point River basin; (a) Observation (b) Forecasts at Lead Time 1H; (c) Forecasts at Lead Time 6H; (d) Forecasts at Lead Time 12H; (e) Forecasts at Lead Time 18H.

The time-series of observed domain-averaged rainfall (figure 4(a)) shows a major rain event (during the period of hours 24 to 70) that was characterized by a gradually varying rainfall accumulation. Figure 3.4 shows overall good agreement between the time series of domain-averaged precipitation observation and forecasts. The 12-h and 18-h lead forecasts tended to underestimate precipitation during some hours (e.g., hours 50–60), and the 1-h lead forecasted tended to overestimate in some hours (e.g., hours 53–57).

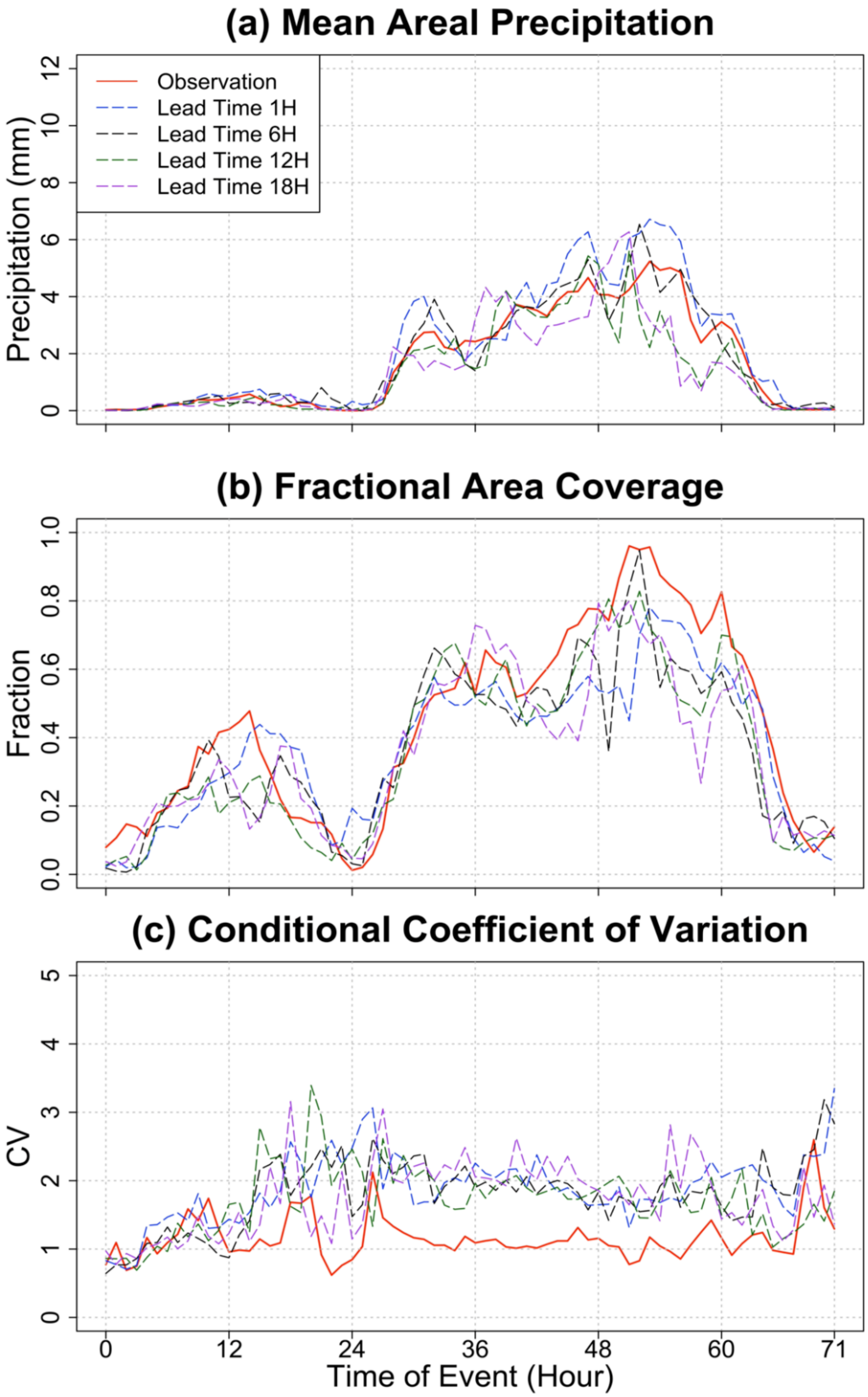


Figure 3.4. Event overview of frontal system driven storms in Southern Midwestern US: (a) Time series of mean areal precipitation; (b) Time series of fractional area coverage; (c) Time series of conditional coefficient of variation.

The time series of frain (figure 4(a)) was bimodal and differed from the unimodal pattern of domain-averaged precipitation. Apparently, the light rain events during the first 24 h (which are not noticeable in the domain- averaged precipitation time series) covered large areas of the domain (up to 50%) causing one of the peaks in the frain time series. The major rainy period (from hour 34 to 65), which covered 60% to 100% of the domain, caused the second peak. Therefore, both light and heavy rain events covered large areas of the domain. How did the forecasts saw this bimodal pattern? All the forecasts captured well this bimodal pattern, but tended to underestimate the actual magnitudes of frain during both peaks.

Figure 3.4c shows that the conditional coefficient of variation of observed rainfall fields is mostly constant around 1, however, the forecasts produced a value of 2, indicating that the forecasts produced more spatially variable rainfall fields (doubling the spatial standard deviation) than what the observations show. The higher spatial variability of the forecasted fields compared to the observations is not surprising as the forecasts produce relatively large number of scattered high-intensity rainfall fields (see Figures 3.3(b)–(e)).

3.2.3 The August 25-31, 2017, Hurricane Harvey Storms in Texas

Hurricane Harvey was the most significant cyclone rainfall event in United States history (Blake and Zelinsky 2017). On August 25th, Hurricane Harvey made landfall near Rockport, Texas. The forward motion of Hurricane Harvey slowed down as it moved inland, producing tremendous rainfall amounts in southeastern Texas and southwestern Louisiana (Watson et al 2018). Figure 3.5. shows the spatial map of accumulated rainfall according to observations. Hurricane Harvey brought very large amount of rainfall to the region, with rainfall accumulation exceeding 1,000 mm in some large areas around the center of the domain. The 6-h, 12-h, and 18-h lead forecasted captured well the spatial pattern of observed precipitation, while the 1-h lead forecast misplaced some part of the storm further north.

The time-series of domain-averaged accumulated precipitation is shown in figure 6(a). The hurricane brought strong storms through the entire storm duration, about five days. There was a high degree of agreement between the temporal pattern of forecasts and observations. In terms of actual magnitudes, all the forecasts produced mean-averaged values that are closer to the observations, except the 1-h lead forecast which resulted in overestimation throughout the entire storm period.

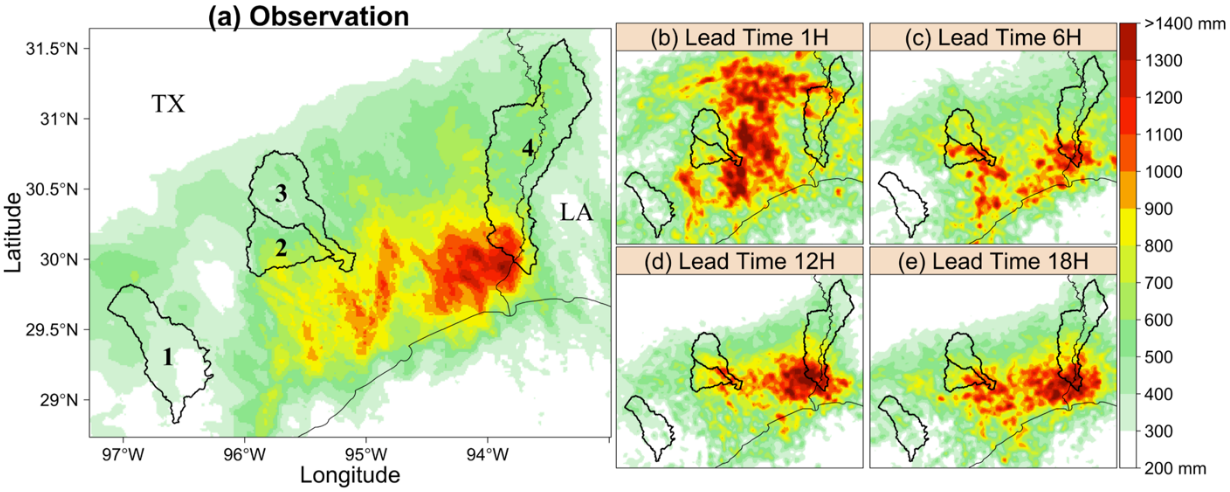


Figure 3.5. Precipitation accumulation map (7-days) of Hurricane Harvey driven storms in Texas: (1) Navidad River basin; (2) Cypress Creek basin; (3) West Fork San Jacinto River basin; (4) Lower Sabine River basin; (a) Observation (b) Forecasts at Lead Time 1H; (c) Forecasts at Lead Time 6H; (d) Forecasts at Lead Time 12H; (e) Forecasts at Lead Time 18H.

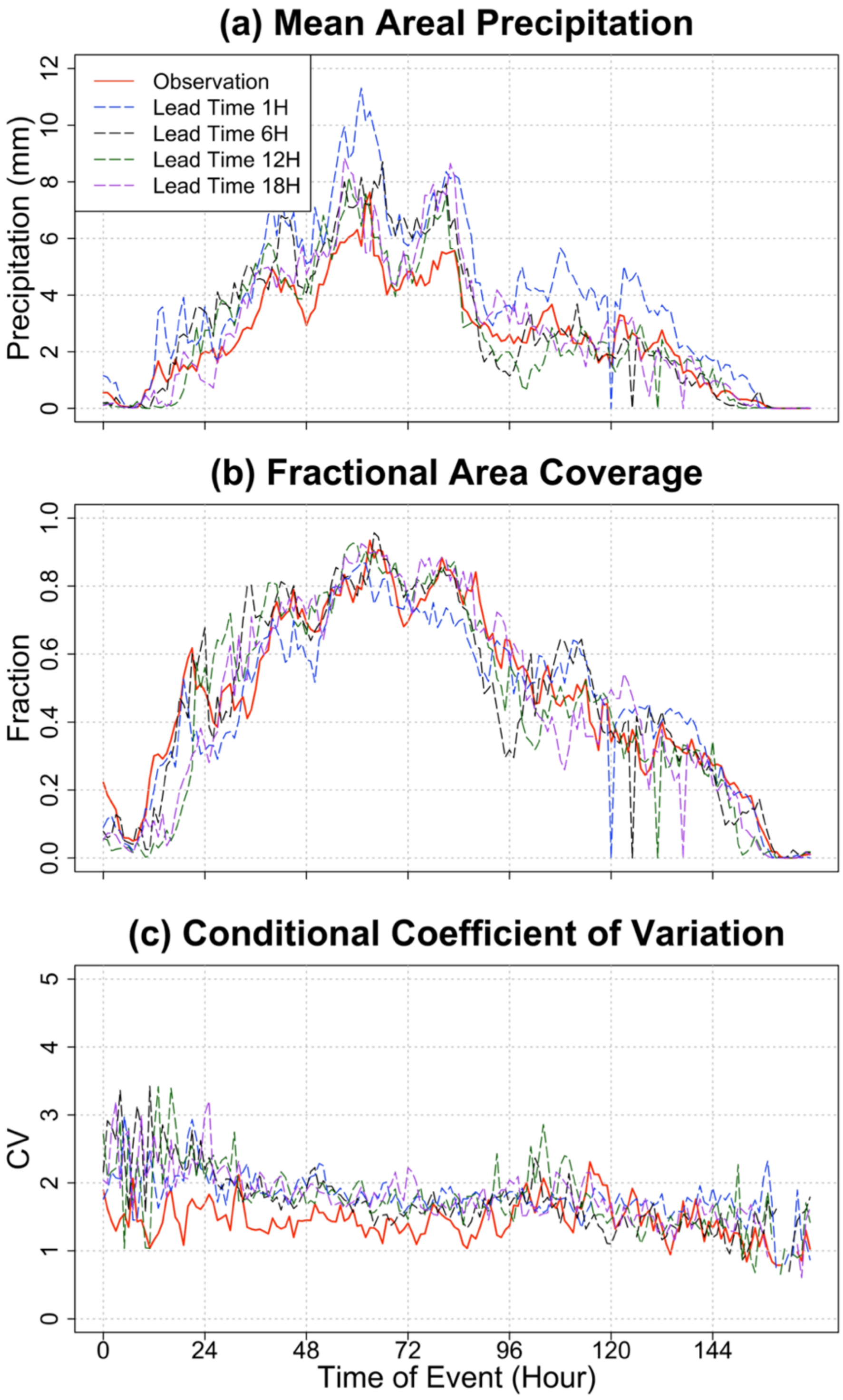


Figure 3.6. Event overview of Hurricane Harvey driven storms in Southern Midwestern US: (a) Time series of mean areal precipitation; (b) Time series of fractional area coverage; (c) Time series of conditional coefficient of variation.

Figure 3.6b shows that the time series of frain had a wide bell-shaped pattern, showing significant areas of the domain covered by rain at all times. There was a high degree of agreement between the forecasts (including the 1-hr forecast) and observations in terms of hourly time series of fraction of domain covered by rain throughout the storm duration.

According to Figure 3.6c, the conditional coefficient of variation of observed rainfall fluctuates within 1 and 2, and does not show association with magnitudes domain-averaged rainfall. The CVcon values for the forecasts were closer to those for the observed values: for almost half of the storm duration, the forecasted CVcon values were slightly higher than the observed values (meaning the rainfall fields are spatially more variable), while for the other half of the storm hours the forecasted values were similar to the observed values.

3.2.4 The September 13-17, 2018, Hurricane Florence Storms in the Carolinas

Hurricane Florence was the first major hurricane of the 2018 Atlantic hurricane season. On September 14th, 2018, Hurricane Florence moved inland, and the slowing down forward motion produced large amounts of rainfall across the Carolinas (Feaster et al 2018). Figure 3.7 shows that Hurricane Florence brings large amount of rainfall to the eastern part of the domain with very large amounts (>700 mm) in some localized areas. All the forecasts captured this rainfall pattern, but produced larger areas with very high rainfall amounts than what the observations show. The overestimation by the rainfall forecasts holds for all the lead-time forecasts.

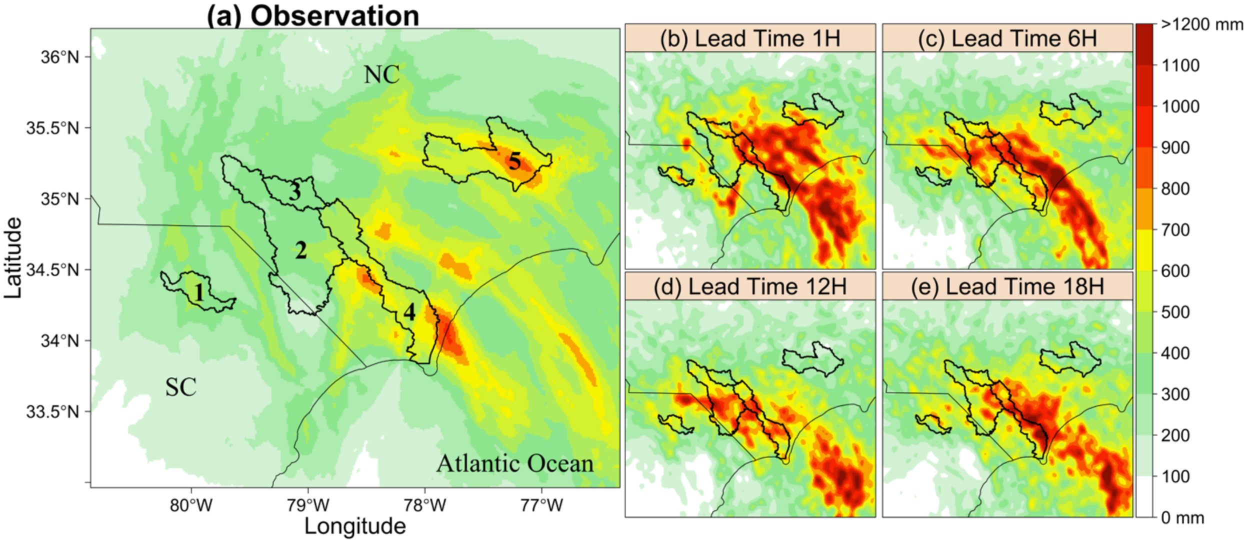
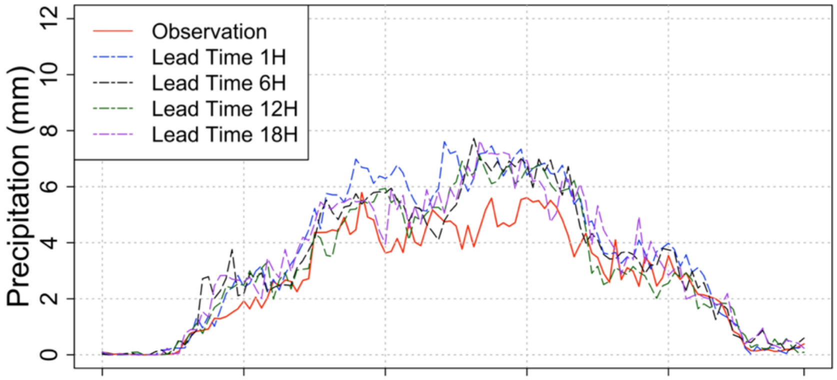


Figure 3.7. Precipitation accumulation map (5-days) of Hurricane Florence driven storms in North Carolina and South Carolina: (1) Black River basin; (2) Lumber River basin; (3) Rockfish creek basin; (4) Cape Fear River basin; (5) Neuse River basin; (a) Observation (b) Forecasts at Lead Time 1H; (c) Forecasts at Lead Time 6H; (d) Forecasts at Lead Time 12H; (e) Forecasts at Lead Time 18H.

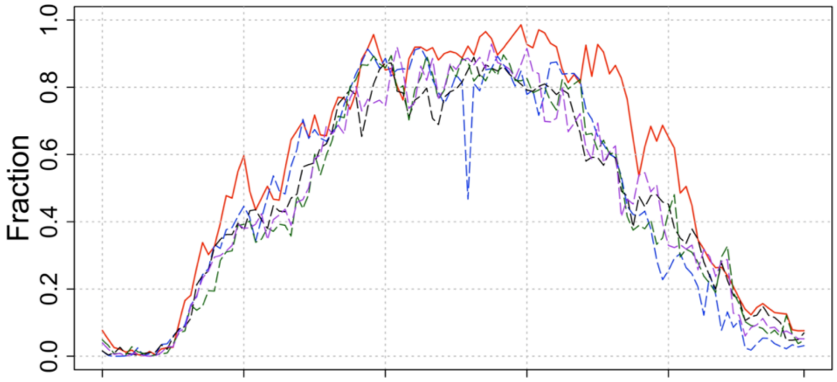
The temporal variability of the domain-averaged precipitation during the storm period was examined in figure 8. The time series rainfall intensity (figure 8(a)) shows that the hurricane brought large amounts of rainfall throughout the storm period, about 4 days, resulting in a wide bell-shaped pattern. There was a good agreement between the temporal variability of forecasts and observations. However, the forecast overestimated precipitation during the peak hours, consistent with the spatial maps discussed above.

As shown in in figure 8(b), the fraction of the domain covered by rain also exhibited pronounced bell-shaped pattern. This pattern was well reproduced by all the forecasts. However, the forecasts tended to slightly underestimate frain throughout the storm period. As displayed in figure 8(c), the conditional coefficient of variation fluctuated within 1.5 ± 1, and showed no association with rainfall amount. The forecasts produced this pattern, but tended to slightly overestimate CV.

(a) Mean Areal Precipitation



(b) Fractional Area Coverage



(c) Conditional Coefficient of Variation

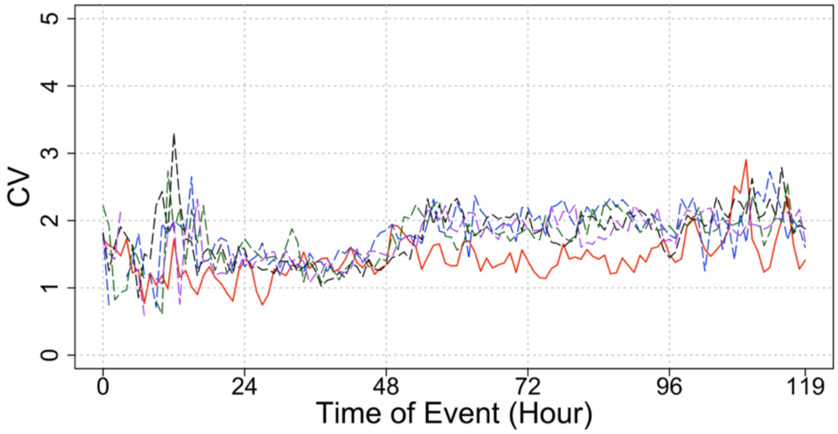


Figure 3.8. Event overview of Hurricane Florence driven storms in North Carolina and South Carolina: (a) Time series of mean areal precipitation; (b) Time series of fractional area coverage; (c) Time series of conditional coefficient of variation.

3.2.5 The September 4-6, 2019, Hurricane Dorian Storms in the Carolinas

Hurricane Dorian was the first major hurricane of the 2019 Atlantic hurricane season and was one of the most powerful hurricanes on record in the open Atlantic region (Cerrai et al., 2019). Instead of landing on the Carolinas, Hurricane Dorian swiped the coastal area of the North Carolina and South Carolina, bringing moderate precipitation. The observed rainfall covers the central and eastern parts of the region, with higher rainfall values forming stripes (Figure 3.9). All the forecasts detected the occurrence and severity of rainfall, however, there were differences in the spatial location of high rainfall values, with the forecasts producing more scattered patterns.

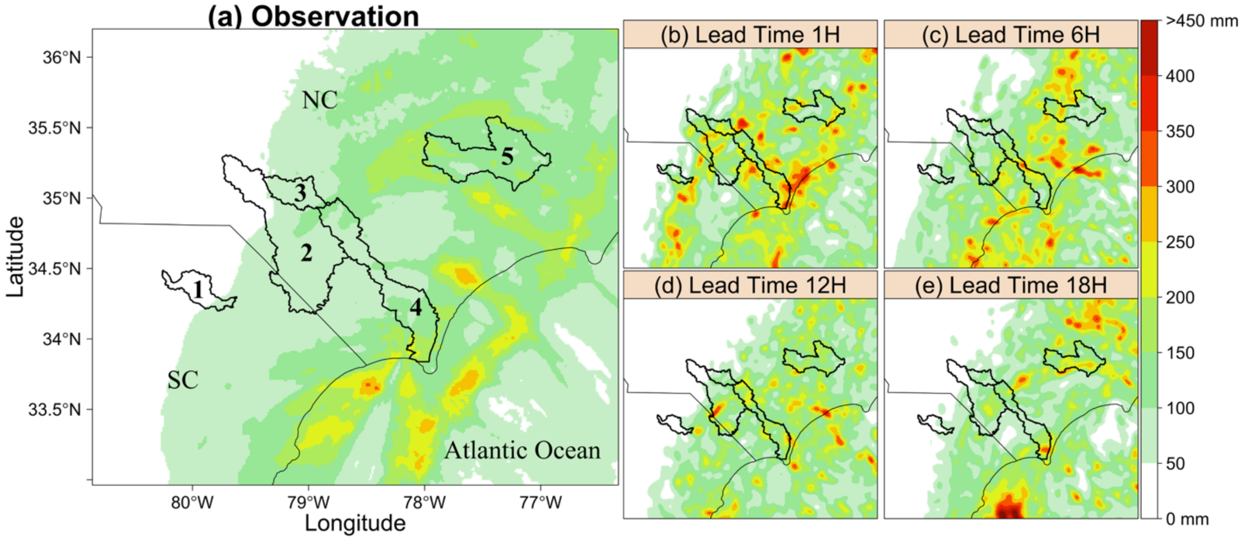
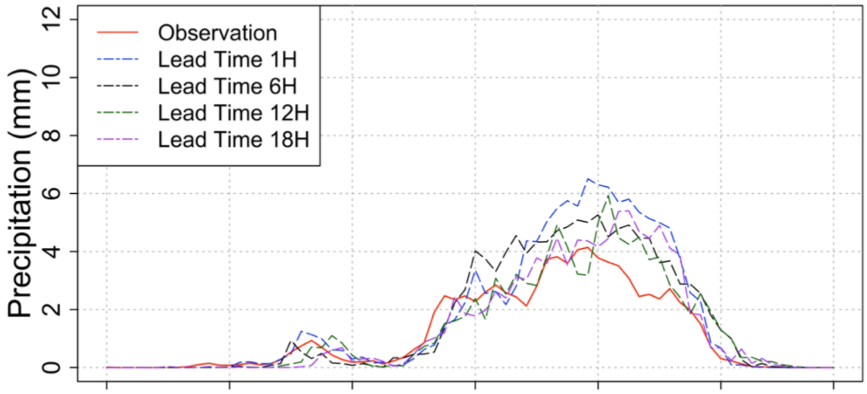


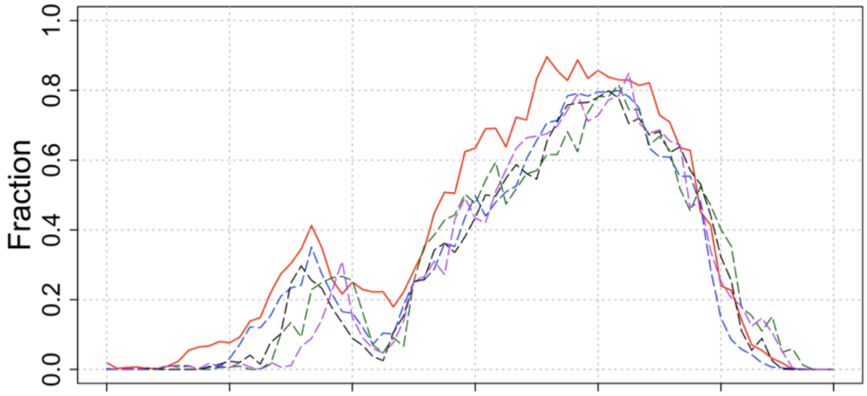
Figure 3.9. Precipitation accumulation map (3-days) of Hurricane Dorian driven storms in North Carolina and South Carolina: (1) Black River basin; (2) Lumber River basin; (3) Rockfish creek basin; (4) Cape Fear River basin; (5) Neuse River basin; (a) Observation (b) Forecasts at Lead Time 1H; (c) Forecasts at Lead Time 6H; (d) Forecasts at Lead Time 12H; (e) Forecasts at Lead Time 18H.

According to the time-series of domain-averaged rainfall (Figure 3.10a), all the forecasts captured well the temporal dynamics of observed rainfall, albeit with some overestimation. While the 1-hr lead forecast tended to have relatively larger bias coverage (Figure 3.10b), the observed rainfall shows a bimodal pattern, with one relatively smaller peak (around hour 20), and another much larger peak (during the period of hours 24 to 70) that was characterized by a gradually varying accumulation. All the forecasts reproduced the temporal dynamics of frain, with slight underestimation bias. Joint analysis of Figure 3.10a and 2.10b reveals that the forecasted fields underestimated the areal coverage of conditional coefficient of variation (Figure 3.10c), the observed rainfall has CV within the range between 1 and 2, whereas the forecasted fields have corresponding values much higher than this indicating that the forecasted fields produced “rougher” spatial structure than observed.

(a) Mean Areal Precipitation



(b) Fractional Area Coverage



(c) Conditional Coefficient of Variation

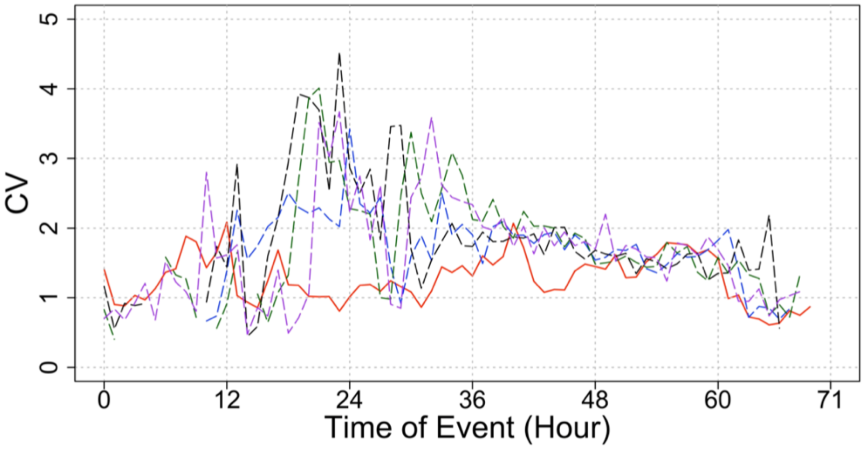


Figure 3.10. Event overview of Hurricane Dorian driven storms in North Carolina and South Carolina: (a) Time series of mean areal precipitation; (b) Time series of fractional area coverage; (c) Time series of conditional coefficient of variation.

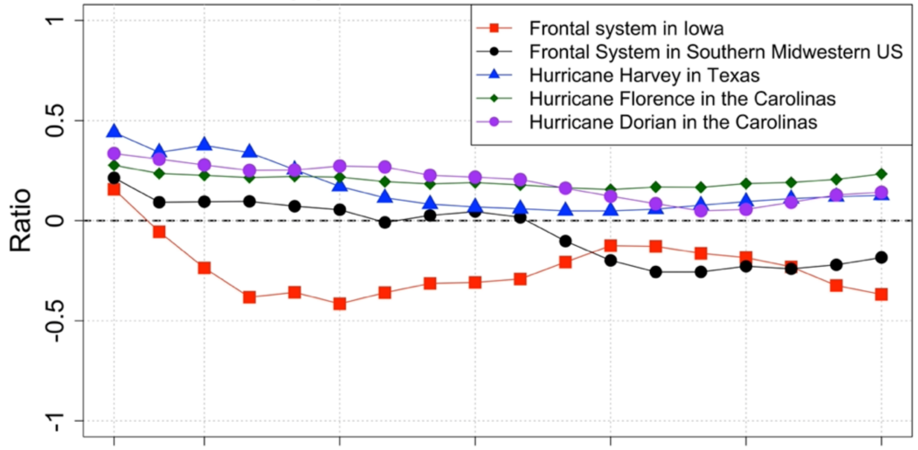
3.2.6 Evaluation of Precipitation Geometric Features

Here, we evaluate the ability of the forecasts to reproduce the geometric features of the accumulated precipitation, in terms of the Amplitude, Structure, and Location components. Figure 3.11 shows the ASL results for each storm, as a function of forecast lead time. The Amplitude values (Figure 3.11a) for both hurricanes are positive at all lead times, indicating overestimation of domain-averaged rainfall. The Amplitude for Hurricane Florence is constant around 0.20 at all lead times, indicating that the forecasts overestimate the domain-averaged rainfall by a factor of 0.6. Similarly, the Amplitude of Hurricane Dorian is constant around 0.3 from lead time 1h to 12h but drops to 0.1 to 0.2 thereafter. The Amplitude of Hurricane Harvey shows some variation with respect to lead time: it is generally higher at short lead times (1 h to 6 h), gets close to zero between 8 h and 13 h lead times, and slightly increases beyond 13 h. The Amplitude of Frontal system in Iowa is negative at all lead times indicating underestimation of domain-averaged rainfall, except for the 1-h lead time forecast which had a positive amplitude. For this storm, the negative amplitudes are generally higher at lead times from 4 h to 10 h (reaching amplitude of -0.5 at 6-h lead time, corresponding to underestimation of 1.5), and slightly improve beyond the 10-h lead forecast. The Amplitude of Frontal system in Southern Midwestern US was slightly positive during short lead times (1 h – 10 h) but showed larger negative amplitude for lead times exceeding 10 h.

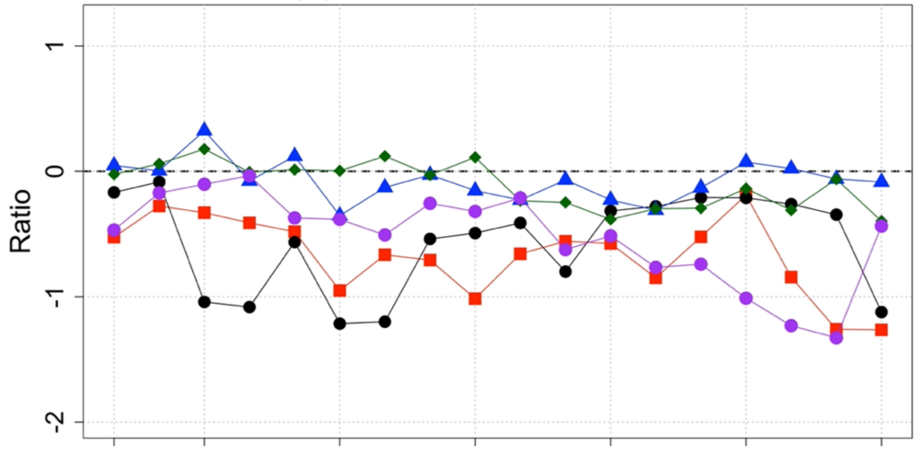
The Structure (Figure 3.11b) values for Hurricane Harvey and Hurricane Florence are close to zero, indicating that the forecasted rainfall fields have comparable spatial structure with the observed rainfall fields. The Structure values for Hurricane Dorian are slightly larger and negative up to lead time of 10 hours, and get larger as the lead time increased further. However, the Structure values for both frontal storms are relatively large and negative, indicating that the forecasted rainfall fields contain too small precipitation objects or too peaked objects, or a combination of these factors, compared to the observed rainfall fields.

The Location (Figure 3.11c) values are very large for the frontal storm in Iowa, for lead times exceeding 5 h, indicating large misplacement of storm location. For lead times under 5 h, the frontal storm in Iowa had smaller storm location error. The Location values for Hurricane Harvey are the smallest at all lead times, indicating that the forecasts reproduced well the storm location. Hurricane Florence and Hurricane Dorian have constant but moderate locational value (around 0.18) at all lead times. The Frontal storm in Southern Midwestern US had moderate locational errors which depend on the lead time: the locational errors are relatively high at short lead times (1 h – 7h), get low for lead times of 8 h – 13 h, and increase for lead times of 14 h and beyond.

(a) Amplitude Component



(b) Structure Component



(c) Location Component

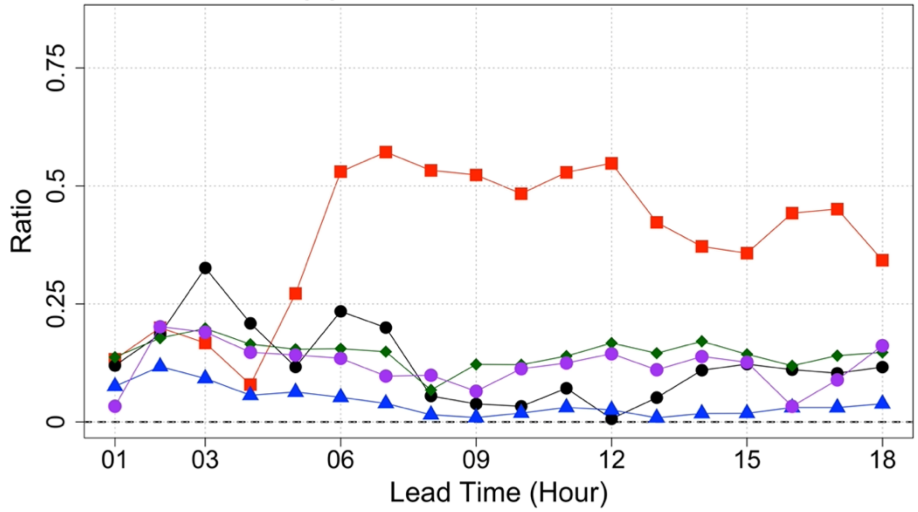


Figure 3.11. SAL analysis results of five selected extreme events at all lead times (1-18H): (a) Amplitude component; (b) Structure component; (c) Location component.

3.2.7 Effect of Spatial Scale on Bias Estimation

In the above sections, we quantified the bias in rainfall forecasts averaged over a large domain. In hydrological simulations, where HRRR forecasts are used as input into a hydrologic model, estimates of bias are required over smaller spatial scales commensurate with the model computational grid. Here, we quantify the bias in the forecasts at smaller spatial scales, from 2 km × 2 km to increasingly coarser spatial scales all the way up to 128 km × 128 km. We perform sampling experiment, where we divide the domain into grids according to the spatial scale of interest and compare the grid-averaged forecast with the corresponding grid-averaged observed rainfall.

Figure 3.12 presents the relative bias (defined as: (forecasted rainfall – observed rainfall)/observed rainfall)) distribution for each of the storms. The relative bias at the 2 km × 2 km grids mostly varies in the range -1 to +1 (meaning up to 100% overestimation and up to 100% underestimation). Therefore, the bias estimates obtained at large (hundreds of km grids) are not transferable to bias estimates at smaller grids, indicating that local bias correction are more preferable over global bias correction methods. In addition, the bias estimates at 2 km × 2 km grids are highly variable from one grid to another, indicating that the bias estimates at one grid is not transferable to another grid even within the same region. The relative bias does not show major reduction as the grid size increases.

To further examine the spread of the relative bias, we selected a few real watersheds from each domain, and calculated the relative bias for each watershed. The relative bias results for the 6-hr lead forecast are shown in Table 3.2. For the frontal storm in Southern Midwestern US, while the bias for the entire domain is about +6%, the bias for the watersheds varies from -12% for Illinois (4,283 km2) to +47% for Flint (328 km2). For the hurricane in Texas, the bias for the entire domain is +19%, while the bias for the watershed varies from -24% for Navidad (3,628 km2) to +56% for WF San Jacinto (2,801 km2). For Hurricane Florence, the bias for the entire domain is +24%, while the bias ranges over a large range: -17% for Neuse (2,759 km2) to +110% for Lumber (4,540 km2). For Hurricane Dorian in the same region, the relative bias for the entire domain is about ±32%, while the bias of individual watersheds ranges from ±12% (Neuse) to ±81% (Black). These results underline that the bias estimates have large spatial variations, and therefore large-scale bias estimates are not transferable to smaller scales.

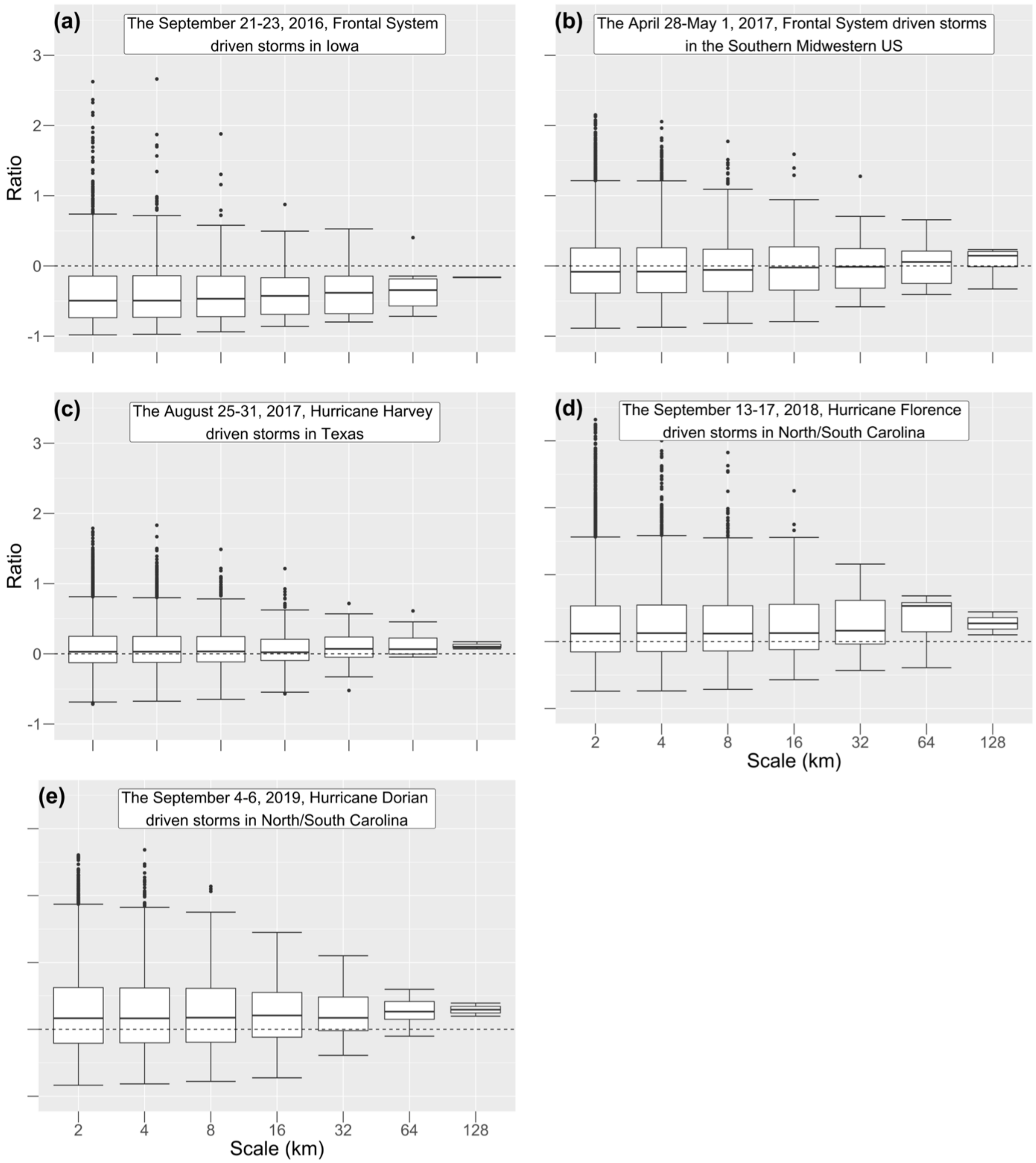


Figure 3.12. Relative bias calculated from precipitation accumulation maps at different scales for the five extreme events: (a) Frontal System driven storms in Iowa; (b) Frontal System driven storms in Southern Midwestern US; (c) Hurricane Harvey driven storms in Texas; (d) Hurricane Florence driven storms in South/North Carolina; and (e) Hurricane Dorian driven storms in South/North Carolina.

Table 3.2. Relative bias at Lead Time 6H calculated over full domain and specific watersheds.

	Watershed Area (km^2)	Relative Bias (Ratio)
The September 21-23, 2016, frontal storms in Iowa		
Full Domain	49140	-0.344
Cedar	17434	-0.438
The April 28-May 1, 2017, frontal storms in the Southern Midwestern US		
Full Domain	176022	0.056
Meramec	5568	0.189
Illinois	4283	-0.116
Eleven Point	3115	0.380
Gasconade	2676	0.261
Big	2513	0.342
Flint	328	0.469
The August 25-31, 2017, Hurricane Harvey storms in Texas		
Full Domain	127575	0.187
Lower Sabine	6844	0.085
Navidad	3628	-0.243
WF San Jacinto	2801	0.560
Cypress	1960	0.486
The September 13-17, 2018, Hurricane Florence storms in the Carolinas		
Full Domain	99828	0.244
Lumber	4540	1.108
Cape Fear	2811	0.446
Neuse	2759	-0.174
Rockfish	801	0.944
Black	757	0.020
The September 4-6, 2019, Hurricane Dorian storms in the Carolinas		
Full Domain	99828	0.316
Lumber	4540	0.554
Cape Fear	2811	0.480
Neuse	2759	0.117
Rockfish	801	0.170
Black	757	0.812

3.3 Summary and Conclusion

This study evaluated the accuracy of short-range (lead times ranging from 1 h to 18 h) forecasts, for five extreme events in the United States which covered two frontal storms and two hurricanes: the September 21-23, 2016, frontal storms in Iowa, (2) the April 28-May 1, 2017, frontal storms in the Southern Midwestern US, (3) the August 25-31, 2017, Hurricane Harvey storms in Texas, (4) the September 13-17, 2018, Hurricane Florence storms, and (5) the September 4-6, 2019, Hurricane Dorian storms in the Carolinas. The basis of the investigation was the HRRR operational forecasts, which are used as input into the National Weather Service' National Water Model (NWM). Evaluation of the forecasts was carried out by comparison with high-quality and independent rainfall observational products known as the gauge-corrected Multi-Radar/Multi-Sensor (MRMS-GC). The main results are summarized as follows.

First, the temporal variability of precipitation during the storm period was examined. There was a good agreement between area-averaged forecasts and observations, on an hourly scale. However, the forecasts were mostly biased. The forecasts tend to overestimate rainfall for both hurricanes. However, the forecasts tend to underestimate the frontal storm in Iowa but produced almost unbiased estimates for the Southern Midwestern US.

Examination of the spatial precipitation pattern was additionally carried out. The forecasts were able to capture the spatial pattern of hurricanes, albeit with overestimation. However, the forecasts produced too many, localized, high-rain intensities for the frontal storms. In addition, the forecasts have difficulty locating the single supercell for the frontal storm in Iowa.

With regard to the effect of lead time, the 1 h lead forecast had lower accuracy compared to the other lead-time forecasts. For lead times ranging from 2 h to 18 h, there was not much systematic difference in accuracy among the various lead-time forecasts.

The bias in the forecast were also examined at different spatial scales, ranging from 2 km × 2 km all the way to 128 km × 128 km. The bias estimates for the small spatial scale varied quite a lot, mostly within the range of -100% to +100%, indicating that the bias estimates obtained at large scale (hundreds of km grids) are not applicable to bias estimates at smaller spatial scales, and vice versa. The bias did not also show significant reduction as the rainfall averaging grid increases from 2 km × 2 km all the way to 32 km × 32 km.

In conclusion, the results of our investigation show that the forecasts captured well the temporal variability of observed precipitation, indicating that the HRRR forecasts provide relatively reliable forecasts. In comparison, the forecasts have better accuracy for predicting hurricanes compared to frontal storms, particularly those frontal storms with single super cells. Our results also show that the 1-h lead forecasts showed generally lower accuracy than the other lead-time forecasts.

Finally, we point out that, although the selected storm cases are interesting from meteorological perspective, they are small in number. Thus, the findings of this study can only provide a first insight into the accuracy of HRRR forecasts for extreme precipitation. Additional analysis involving more storm cases and mechanistic approaches will have to be carried out in order to generalize the results.

CHAPTER 4

Evaluation of Medium-Range Forecasts (GFS) over Nile

This chapter has been accepted in the Journal of Hydrometeorology. © American Meteorological Society. Used with permission. This preliminary version has been accepted for publication in the Journal of Hydrometeorology and may be fully cited. The final typeset copyedited article will replace the EOR when it is published.

Yue, H., Gebremichael, M. and Nourani, V., 2021. Evaluation of Global Forecast System (GFS) Medium-Range Precipitation Forecasts in the Nile River Basin. *Journal of Hydrometeorology*. accepted. doi: 10.1175/JHM-D-21-0110.1

4.1 Study Region

The Nile River basin is the longest river system in the world, flowing approximately 6700 km through 11 countries before reaching the Mediterranean Sea. The Nile basin is also known for its complex climatology, ecology and hydro-politics. The climate of the Nile basin varies with latitude (32° N to 4° S) and altitude (0 m to 2800 m MSL), and therefore, the basin has a high climate diversity. According to Köppen climate classification, the Nile basin's climate can be classified from the north to the south as arid, temperate, and tropical. Rainfall in Nile basin has a different pattern of seasonality in different parts of the basin which generally follows the movement of the Intertropical Convergence Zone (ITCZ) (Di Baldassarre et al. 2011; NBI 2012a). Most of the Nile River Basin (NRB) has only one rainy season, except for the equatorial region which has two rainy periods (NBI 2012b). The annual rainfall amount varies from over 2000 mm in Blue Nile sub-basin (Ethiopian highlands) to less than 50 mm in the arid region of Sudan and Egypt in the north (Sutcliffe and Parks, 1999; NBI 2012b; Onyutha 2016).

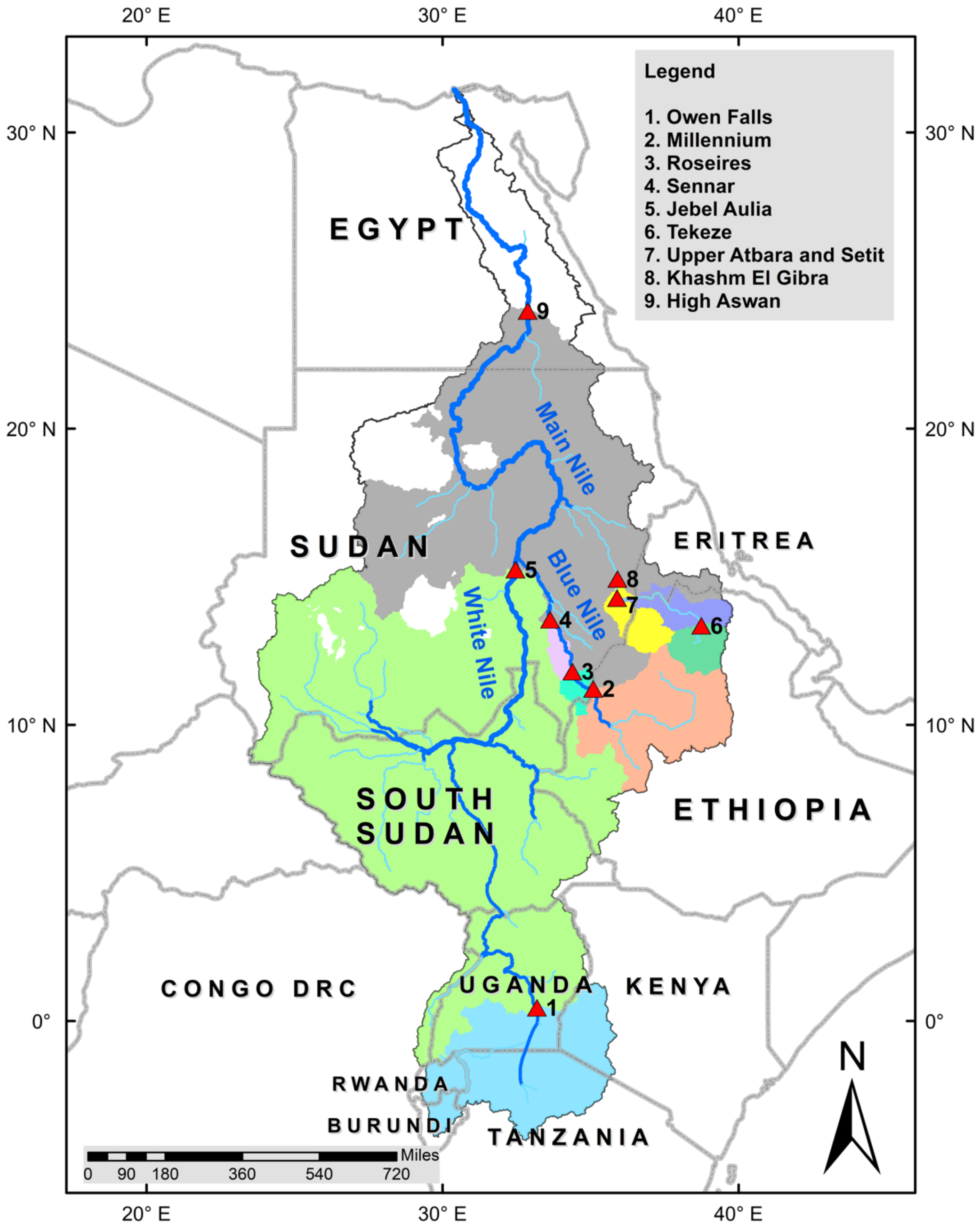


Figure 4.1. The Nile River Basin, showing the sub-basins delineated on the basis of major reservoir dam locations: (1) Owen Falls, (2) Millennium, (3) Roseires, (4) Sennar, (5) Jebel Aulia, (6) Tekeze, (7) Upper Atbara and Setit, (8) Khashm El and (9) High Aswan.

The Nile basin is home to nine major dam reservoirs: High Aswan Dam, Owen Falls Dam, Sennar Dam, Jebel Aulia Dam, Roseires Dam, Millennium Dam, Tekeze Dam, Khashm El Gibrat, and Upper Atbara and Setit Dam (see Table 4.1 for the salient features of the dams). In order for our results to be applicable to reservoir managers, we divided the Nile basin into sub-basins according to the locations of the dam reservoirs (see Figure 4.1). In order to avoid overlap in sub-basins, we define the sub-basin of a dam as the drainage between the dam itself and the previous dam. For example, the drainage basin of the Roseires Dam does not include the drainage basin of the Millennium Dam. The watershed sizes vary over a large range, from 13,895 sq. km for Sennar Dam to 1,396,413 sq. km for Jebel Aulia Dam.

Table 4.1. Major reservoir dams in the Nile basin

Dams	Country	Operational since*	Capacity (million m^3)*	Power (MW)*	Purpose*			Area of Drainage Basin (km^2)**	Mean Elevation of Drainage Basin (m)**	Number of GPCC*** Gauges
					Irrigation and Water Supply	Flood Control	Hydroelectricity			
Owen Falls	Uganda	1954	80000	180	x		x	264210	1392	51
Millennium	Ethiopia	2020	74000	6450			x	174146	1874	15
Roseires	Sudan	1966	3000	280	x		x	14110	680	0
Sennar	Sudan	1925	930	NA	x			13895	473	3
Jebel Aulia	Sudan	1937	3500	30	x		x	1396413	647	35
Tekeze	Ethiopia	2009	9000	300			x	30723	2088	1
Upper Atbara and Setit	Sudan	2017	2700	135	x		x	36104	1440	1
Khashm El Gibra	Sudan	1964	1300	10	x		x	33426	929	1
High Aswan	Egypt	1970	16200	2100	x	x	x	790245	472	17

* information obtained from the Global Reservoir and Dam Database (Lehner et al. 2011) and Food and Agriculture Organization of the United Nations (FAO)'s Global Information System on Water and Agriculture (AQUASTAT).
** Calculated from HydroSEHDS (Lehner et al. 2008).
***GPCC = Global Precipitation Climatology Project

4.2 Results and Discussion

4.2.1 How Well is the Precipitation Total Forecasted?

Figure 4.2 compares the precipitation total (sub-basin-averaged, and annual) obtained from the 1-day lead-time GFS forecast and different satellite precipitation products for each sub-basin. According to our reference precipitation product (i.e. IMERG Final), the annual (15 June 2019 – 15 June 2020) precipitation varies from 210 to 2117 mm, depending on the sub-basin. The largest watershed-average precipitation (> 1000 mm) is observed in the southern and eastern parts of the basin: 2,117 mm (Owen Falls), 1,424 mm (1151 mm (Khashm El Gibrat), 1044 mm (Roseries). The lowest watershed-averaged precipitation is observed in the northern part: 210 mm (High Aswan).

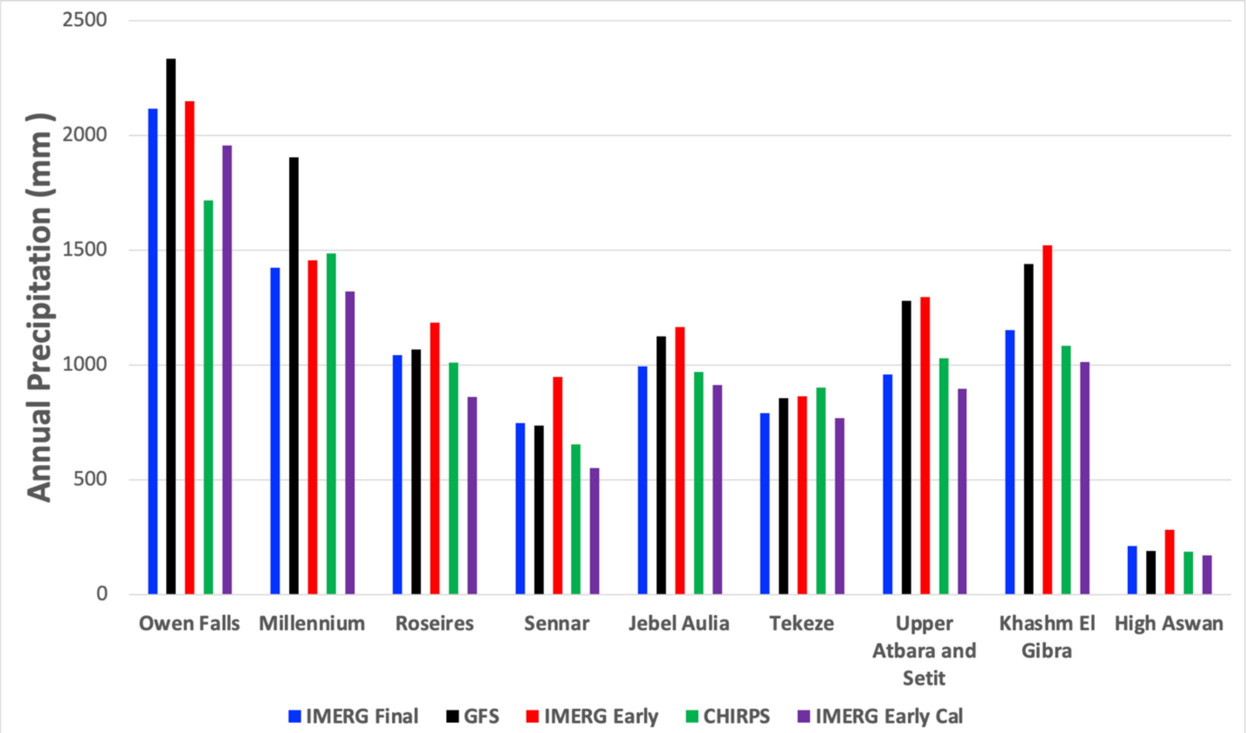


Figure 4.2. Sub-basin averaged annual precipitation (mm) for the period, 15 June 2019 to 15 June 2020, for each of the Nile's sub-basins, derived from the 1-day lead GFS forecast and different satellite precipitation products.

Validated against IMERG Final, GFS has small bias (within 25% of the reference) for six sub-basins: Owen Falls, Roseires, Sennar, Jebel Aulia, Tekeze, and High Aswan. However, GFS shows large overestimation bias for three sub-basins: (Millennium +47% (Millennium), +50% (Upper Atbara and Setit), and +42% (Khashm El Gibrat). The watersheds, where GFS shows large overestimation bias, are characterized by wet climate, location in northern hemisphere, and high-elevation (Millennium) or medium-elevation (Upper Atbara & Setit, and Khashm El Gibrat) areas. The watersheds, where GFS shows relatively smaller bias, are characterized by (1) either wet climate in a southern hemisphere (Owen Falls), or (2) moderate-wet climate in the northern hemisphere (Roseires, Jebel Aulia), or (3) dry climate in the northern hemisphere (Tekeze, Sennar, High Aswan).

Now, let us compare the performance of GFS with the performance of IMERG Early. It is worth noting that for the Millennium watershed (where GFS overestimates by +47%), the IMERG Early estimates have better performance with bias of only +8%. However, for the Upper Atbara & Setit and Khashm El Gibrat watersheds (where GFS overestimates by +50% and +42%), the IMERG Early estimates have similar large overestimation bias of +40% and +31%, respectively. The difference in the relative performance of GFS and IMERG Early over these three watersheds can be partly explained by the accuracy of the reference rainfall, IMERG Final, i.e. by the number of rain gauges used in IMERG Early: there were 15 rain gauges in the Millennium watershed but only 1 rain gauge in each Upper Atbara & Setit and Khashm El Gibrat watershed. However, for the High Aswan watershed (where GFS has a bias of only -4%), IMERG Early estimates have large overestimation bias (+34%), which could not be explained by the number of rain gauges used in IMERG Final, as there were 17 rain gauges used. For the remaining watersheds, both GFS and IMERG Early have comparable performance as measured through the bias. The research-version CHIRPS estimates are within 20% of the IMERG-Final product, at all sub-basins, indicating that the bias estimate for GFS and IMERG Early may not change substantially if the reference product were to change from IMERG Final to CHIRPS.

4.2.2 How Well is the Time Series of Daily Precipitation Forecasted?

Figures 4.3 and 4.4 present time series of basin-averaged daily precipitation derived from the 1-day lead GFS forecasts and various satellite products, focusing on the wet season (i.e. June – September for all sub-basins, except for the Owen Falls sub-basin whose rainy period is October – March). Almost all sub-basins (except High Aswan) experience precipitation on a daily basis during the wet periods. We use the coefficient of variation (CV defined as the standard deviation normalized by the mean) to measure temporal variation.

According to IMERG Final, the largest temporal variability (CV = 1.33 to 1.43) is observed for Roseires, Sennar, and Tekeze sub-basins, and the lowest temporal variability (CV; 0.67 to 0.78) for Millennium, Jebel Aulia, and Owen Falls. Compared to IMERG Final, GFS tends to give smaller variabilities (CVs) at almost all sub-basins but captures well the pattern in CV variation from one sub-basin to another. IMERG Early has CVs that are very close to IMERG Final in most cases, except for Tekeze (where GFS has larger CV) and Roseires (where GFS has lower CV). CHIRPS also has CVs that are very close to IMERG Final in most cases, except for three sub-basins where CHIRPS shows lower CVs (Roseries, Sennar, and Khashm El Gibrat).

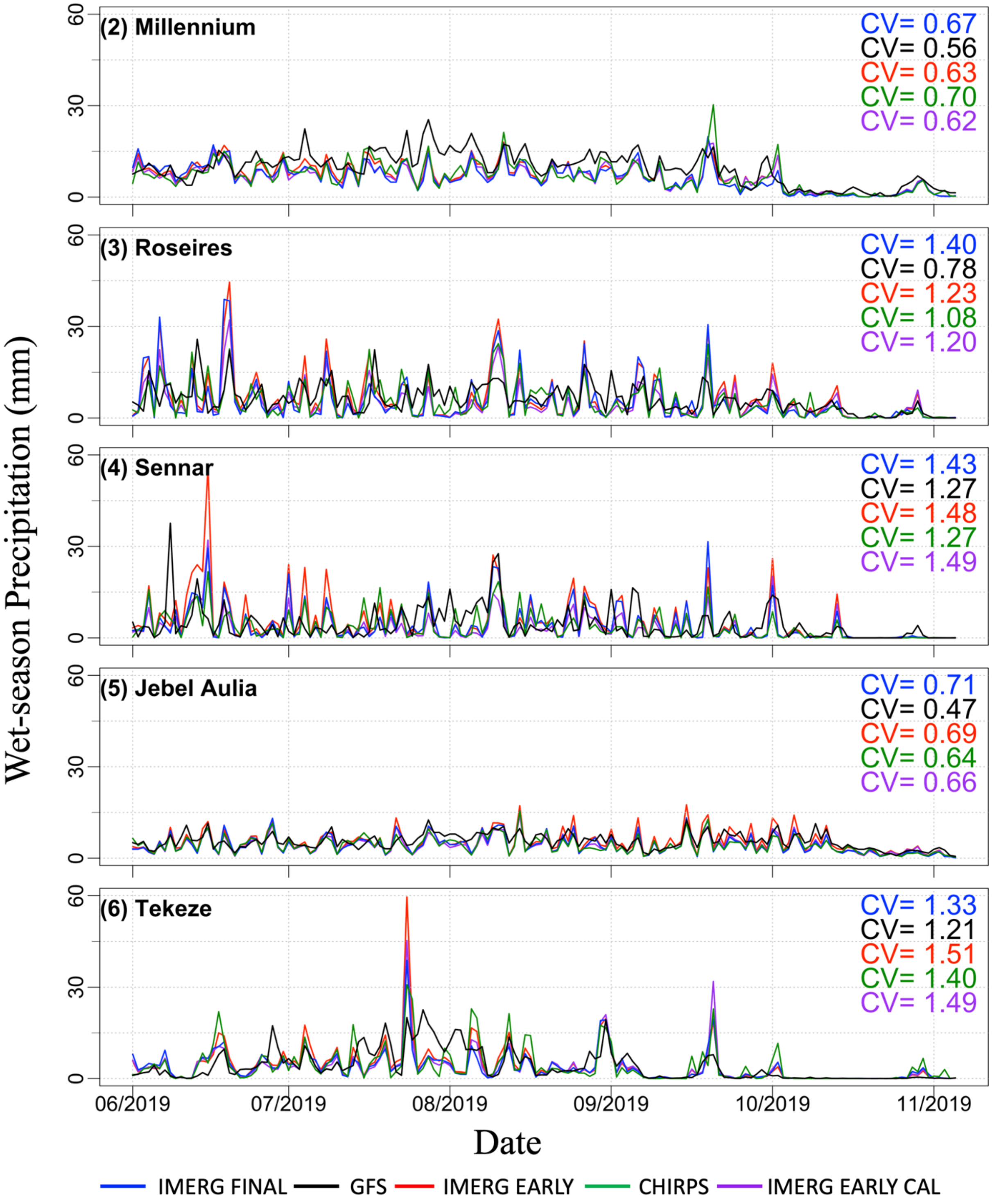


Figure 4.3. Time series of sub-basin averaged precipitation total (mm) for the wet period (June – September 2019 for all sub-basins except for Owen Falls whose wet period is October – March), derived from various precipitation products, for five sub-basins. The Figure also shows the coefficient of variation (CV) as a measure of temporal variation.

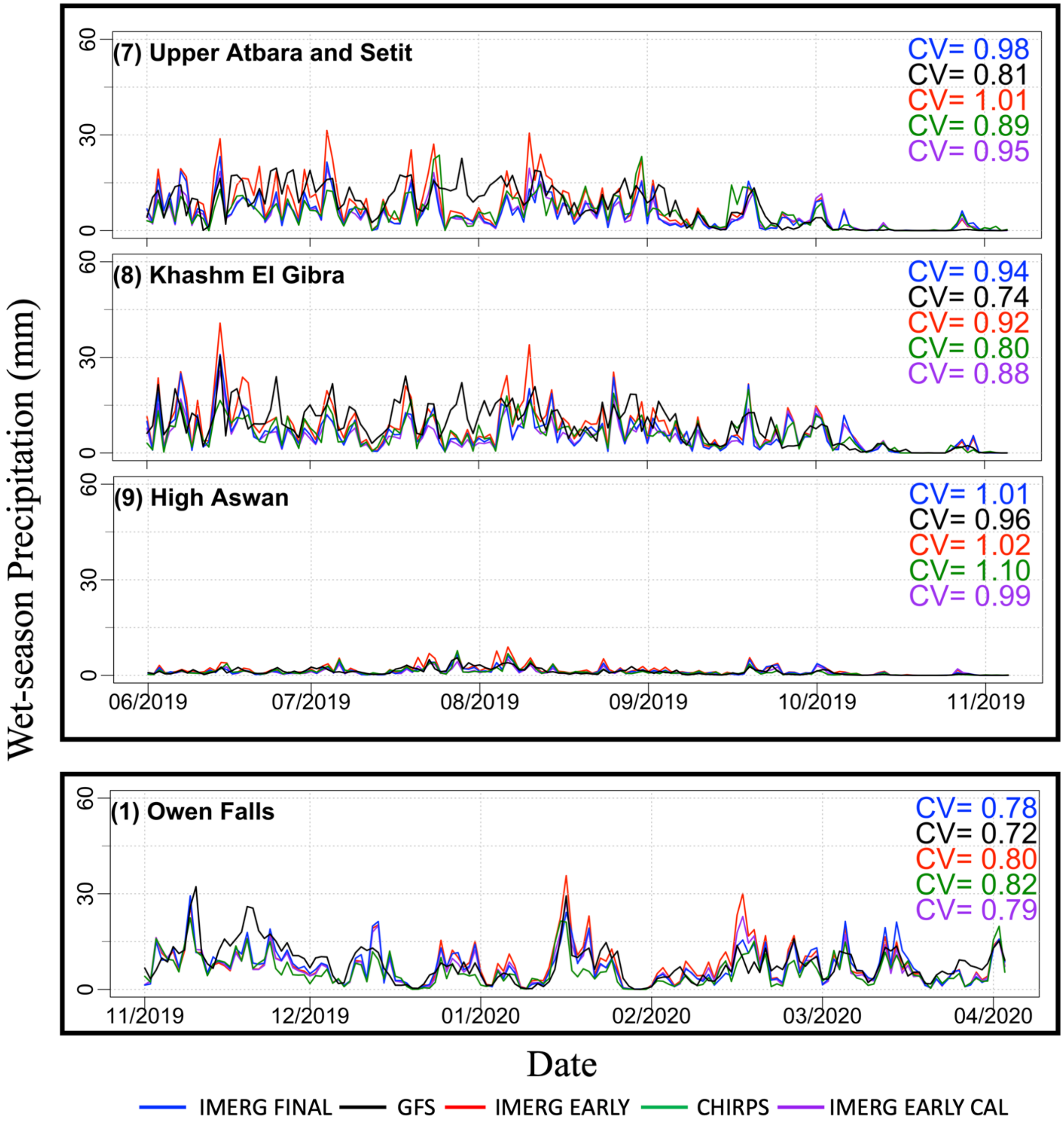


Figure 4.4. Same as Figure 3 but for the remaining four sub-basins.

The performance statistics of various precipitation products (with respect to the IMERG Final estimates) are presented in Figure 4.5, in terms of Kling-Gupta Efficiency (KGE), Bias Ratio (BR), correlation (R), variability ratio (γ), and root mean square error normalized by reference precipitation mean (NRMSE). The KGE scores (Figure 4.5, left panel) for the 1-day lead GFS forecasts are mostly poor ($0 < KGE < 0.5$), except for a couple of sub-basins where KGE scores (Owen Falls, High Aswan) are intermediate ($0.5 < KGE < 0.75$).

The breakdown of the KGE scores (BR, R, and γ) reveals the key factors contributing to the KGE estimates. The GFS tends to overestimate daily precipitation for most sub-basins, as BR is mostly higher than one. The overestimation is particularly higher for Millennium, Upper Atbara & Setit, and Khashm El Gibrat. The correlation coefficient between GFS and IMERG Final varies in the range 0.36 to 0.75. The correlation is higher (> 0.68) for Owen Falls, Millennium, Jebel Aulia, and High Aswan, whereas it is lower (< 0.50) for Roseires and Sennar. Higher correlation values are obtained for large watersheds, and lower correlations for small watersheds, indicating that spatial averaging improves the ability of GFS in capturing the temporal dynamics of watershed-averaged daily rainfall. The variability ratio of GFS is generally lower than IMERG Final, as the ratio varies from 0.56 to 0.98, indicating that the temporal variability of GFS forecasts is general lower compared to those of IMERG Final estimates.

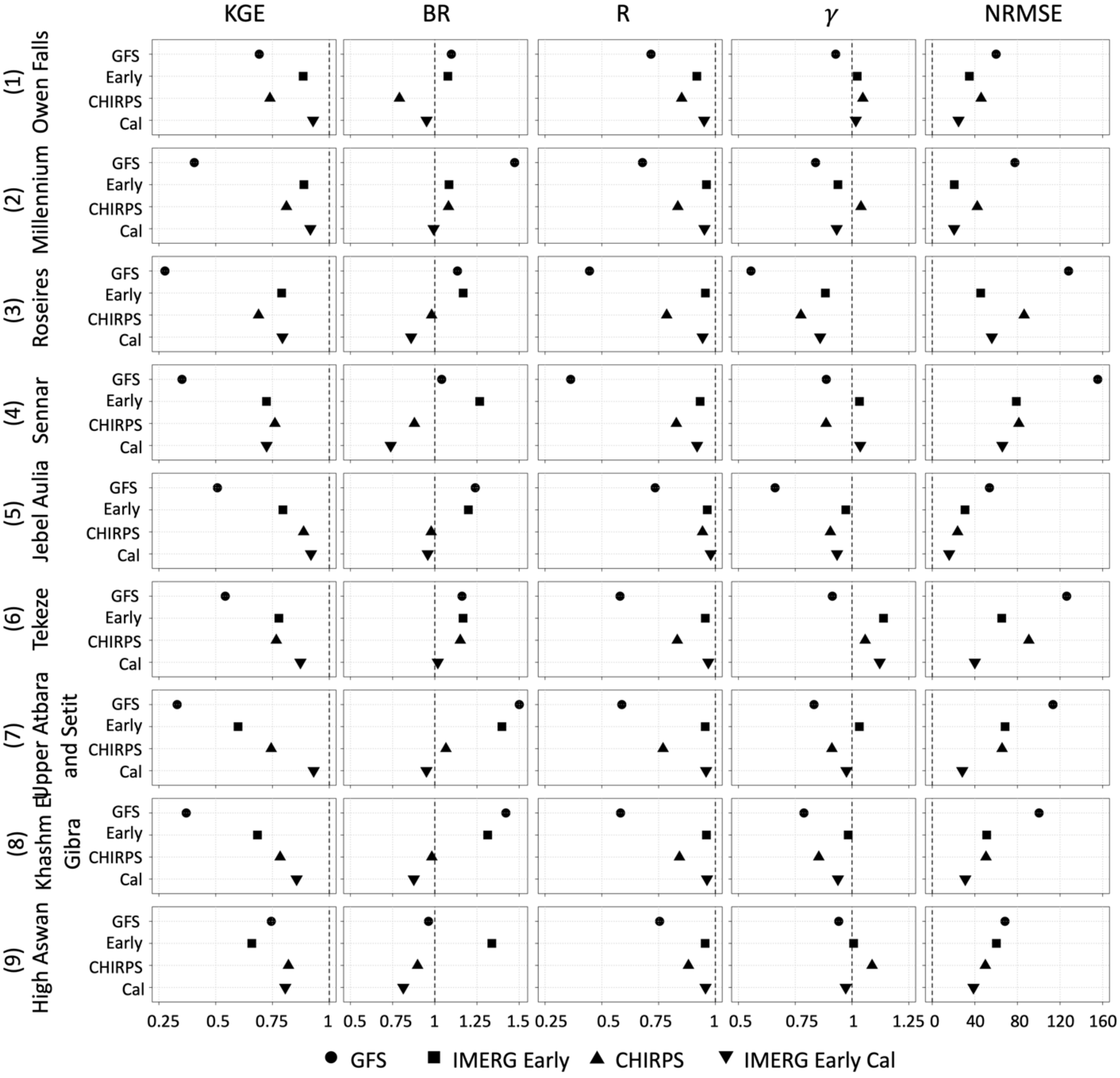


Figure 4.5. Summary of performance statistics (Kling-Gupta Efficiency KGE, Bias Ratio BR, correlation R, variability ratio γ, and root mean square error normalized by reference precipitation mean NRMSE) of 1-day lead GFS forecasts and different satellite products, during the wet periods, at different sub-basins.

The KGE breakdown indicates that the required developments to improve the representation of daily precipitation in GFS forecasts are specific to each sub-basin. For Millennium, Upper Atbara & Setit, and Khashm El Gibrat sites, the low KGE value (0.33 to

0.41) are mainly due to their large overestimation bias. For Roseries, Sennar, and Tekeze the low KGE values (0.28 and 0.35) are due to low correlation and also, in the case of Roseries, low variability. For Jebel Aulia, the lower KGE value (0.51) was mainly due to the low variability. For Roseries, the lower KGE value (0.28) was due to both low correlation and low variability of the GFS. For Owen Falls and High Aswan, the KGE value are relatively better (0.72 and 0.74). High Aswan sub-basin is the driest of all sub-basins, and Owen Falls, located in the southern hemisphere, has a different wet season (October through March) than the remaining sub-basins (June through September).

Next, we examined the performance of IMERG Early with respect to IMERG Final, mainly to assess if we can use the near-real-time IMERG Early product to calibrate and improve the accuracy of GFS forecasts. The overall KGE scores for IMERG Early varies in the range 0.60 to 0.89. IMERG Early tends to overestimate daily precipitation in the range of 8% to 40%, depending on the sub-basin. The overestimation is worse in Upper Atbara and Setit (40%), High Aswan (34%), and Khashm El Gibrat (31%). The correlation and variability ratio between IMERG Early and IMERGE Final is close to one, as one would expect due to the similarity of both products. The main difference between the two products is that unlike IMERG Early, IMERG Final includes monthly rain gauge observations for bias correction. Such monthly-based bias correction procedures would not alter the temporal pattern of IMERG-Final compared to IMERG-Early.

We compared the performances of GFS forecasts and IMERG-Early. In terms of standard error and KGE scores, IMERG Early outperforms GFS in all sub-basins. However, in terms of bias, there is no clear winner between GFS and IMERG-Early. We also compared CHIRPS with IMERG Final to assess how the use of different reference products may affect the finding about the performance of GFS forecasts. The KGE scores of CHIRPS are good (> 0.75) in all cases, indicating that CHIRPS and IMERG Final have comparable KGE performance.

4.2.3 Dependence of Forecast Uncertainty on Precipitation Rate

Here, we assess how the performance of GFS forecasts may vary depending on the magnitude of observed precipitation. Figure 4.6 presents a scatterplot of daily GFS forecasts versus daily IMERG Final estimates. The GFS performance varies from one sub-basin to the other. In Owen Falls and High Aswan, GFS forecasts are almost on the 1:1 line with small spread. In Millennium, Upper Atbara and Setit, and Khashm El Gibrat, GFS overestimates most of the daily precipitation and shows large spread. In Roseires, GFS overestimates precipitation smaller than 10 mm/day but underestimates almost all cases with precipitation higher than 10 mm/day. In Sennar, GFS forecasts have large number of false alarms and severely underestimate precipitation above 10 mm/day. In Jebel Aulia, GFS overestimates precipitation under 5 mm/day but slightly underestimates precipitation above 5 mm/day. In Tekeze, GFS shows large spread. The scatterplot of IMERG Early against IMERG Final is shown in Figure 4.7. IMERG Early estimates are mostly on the 1:1 line with small spread. In sub-basins, where IMERG Early exhibits large overestimation bias (Sennar, Upper Atbara & Setit, and Khashm El Gibrat, and High Aswan), high precipitation rates are overestimated, whereas low precipitation rates are well estimated.

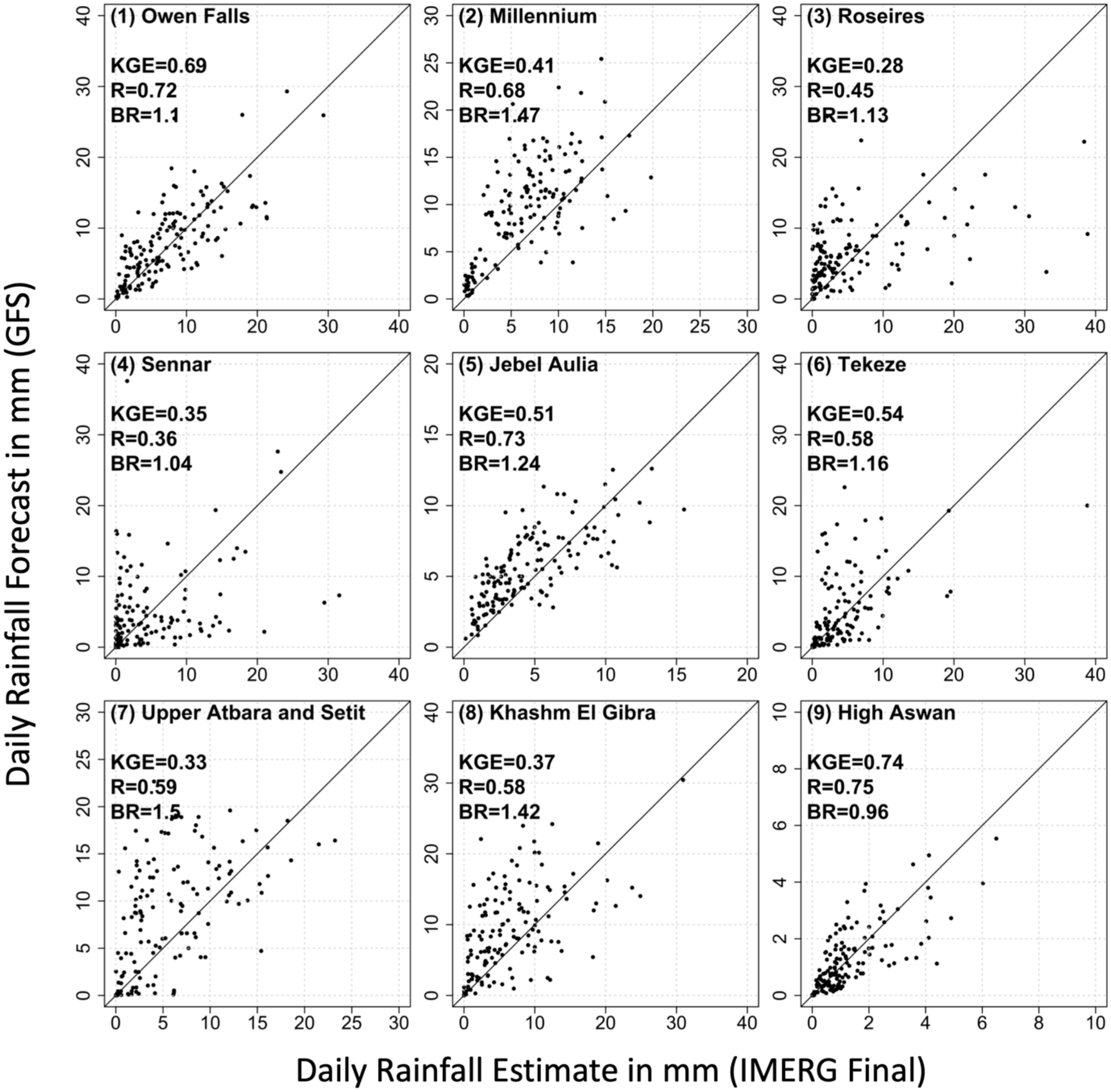


Figure 4.6. Scatterplot of sub-basin averaged daily precipitation forecast obtained from 1-day lead GFS forecast against corresponding values from IMERG Final.

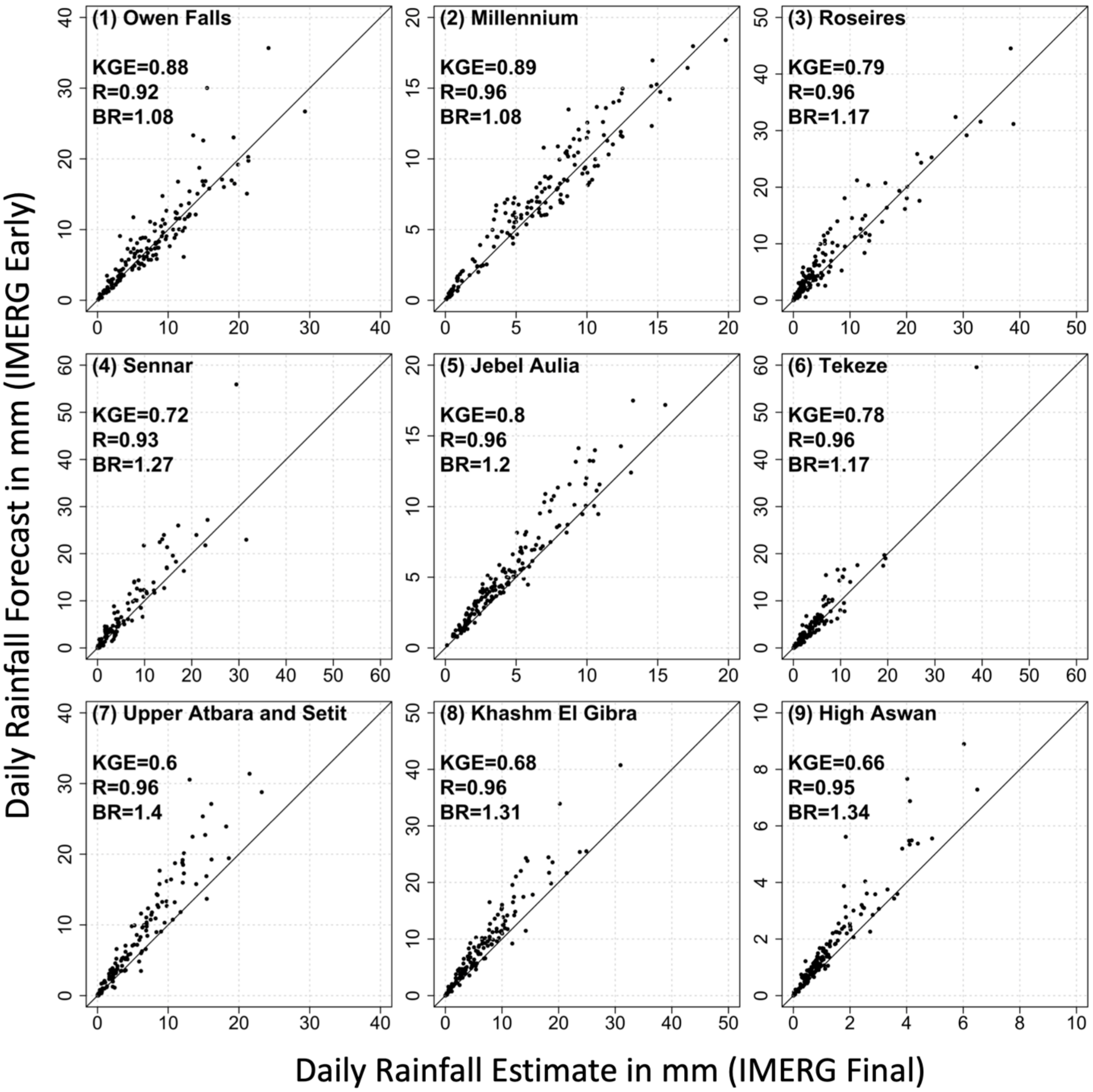


Figure 4.7. Scatterplot of sub-basin averaged daily precipitation forecast obtained from IMERG Early against corresponding values from IMERG Final.

4.2.4 Performance of Climatological Bias Correction of IMERG Early

As discussed above, the GFS watershed-averaged daily precipitation forecasts have poor performance (KGE < 0.5) for seven out of the nine sub-basins, and large overestimation bias for three sub-basins, indicating the need for exploring ways of improving the accuracy of GFS forecasts. One of such ways could be post-processing of GFS forecasts using satellite rainfall observations, particularly near-real time rainfall products, such as, IMERG Early. In order for this approach to be successful, the accuracy of the near-real time rainfall products must be higher than the accuracy of GFS. Comparison of the performance of GFS with that of IMERG Early indicates that while IMERG Early has better bias than the GFS for some basins, it has comparable and worse bias for other basins. This suggests that the bias in the IMERG Early products needs to be reduced such that it can be used for post-processing of GFS forecasts. Here, we explore the possibility of improving the bias in the IMERG Early estimates using a simple climatologically bias-correction method. We performed climatological bias correction of the IMERG Early estimates by calculating the bias ratio between the long-term average (2000-2018) IMERG Final and IMERG Early estimates at each pixel, and multiplying the IMERG Early estimates during the study period (15 June 2019 – 15 June 2020) by the bias ratio at each pixel.

The performance of the climatologically-bias-corrected IMERG Early estimates is included in Figures 4.1 – 4.5. Climatological bias correction has significantly reduced the bias for most sub-basins, with only negligible impact for two sub-basins (Roseries and Sennar). The bias correction has particularly reduced the large bias (from +40% to -5% for Upper Atbara & Setit, from +31% to -12% for Khashm El Gibrat). However, for the two sub-basins (Sennar and High Aswan), where IMERG-Early showed worse bias than the GFS forecasts, the bias-corrected IMERG Early estimates still show worse bias than the GFS, indicating that other bias-correction methods need to be considered.

4.2.5 Dependence of Daily Forecast Uncertainty on Lead Time and Spatial Scale

In order to assess the effect of various lead times and spatial scales on forecast uncertainty, we obtained daily GFS forecasts at various lead times (1-day, 5-day, 10-day, and 15-day), and aggregated the forecasts at spatial scales from 0.25° to coarser scales (0.5°, 0.75°, and 1°) by averaging grids. The purpose of degrading the resolution is to determine at which resolution the forecasts have acceptable performance. The KGE value at each spatial resolution was calculated in the following steps: (i) average the data at the required spatial resolution, (ii) extract pairs of data (one from IMERG Final, and the other from GFS), (iii) concatenate the pairs to form one large series of data, and (4) compute a single KGE from this data series. The resulting KGE values are shown in Figure 4.8 (the KGE components are presented in Supplemental Figures 4.3).

With regard to the effect of spatial scales, the KGE at the GFS native resolution (i.e. 0.25°) is very low, and varies from 0.10 to 0.39 (for 1-day lead), depending on the site. As the spatial scale increases, KGE increases, as expected. For instance, for the Owen Falls site, KGE increases from 0.23 (0.25°, 1-day) to 0.39 (1°, 1-day), and for Tekeze site, KGE increases from 0.35 (0.25°, 1-day) to 0.47 (1°, 1-day). The improvement in KGE with increasing spatial scale comes from improved correlation and variability ratio, as expected due to the effect of averaging (see Supplemental Figures 1- 3).

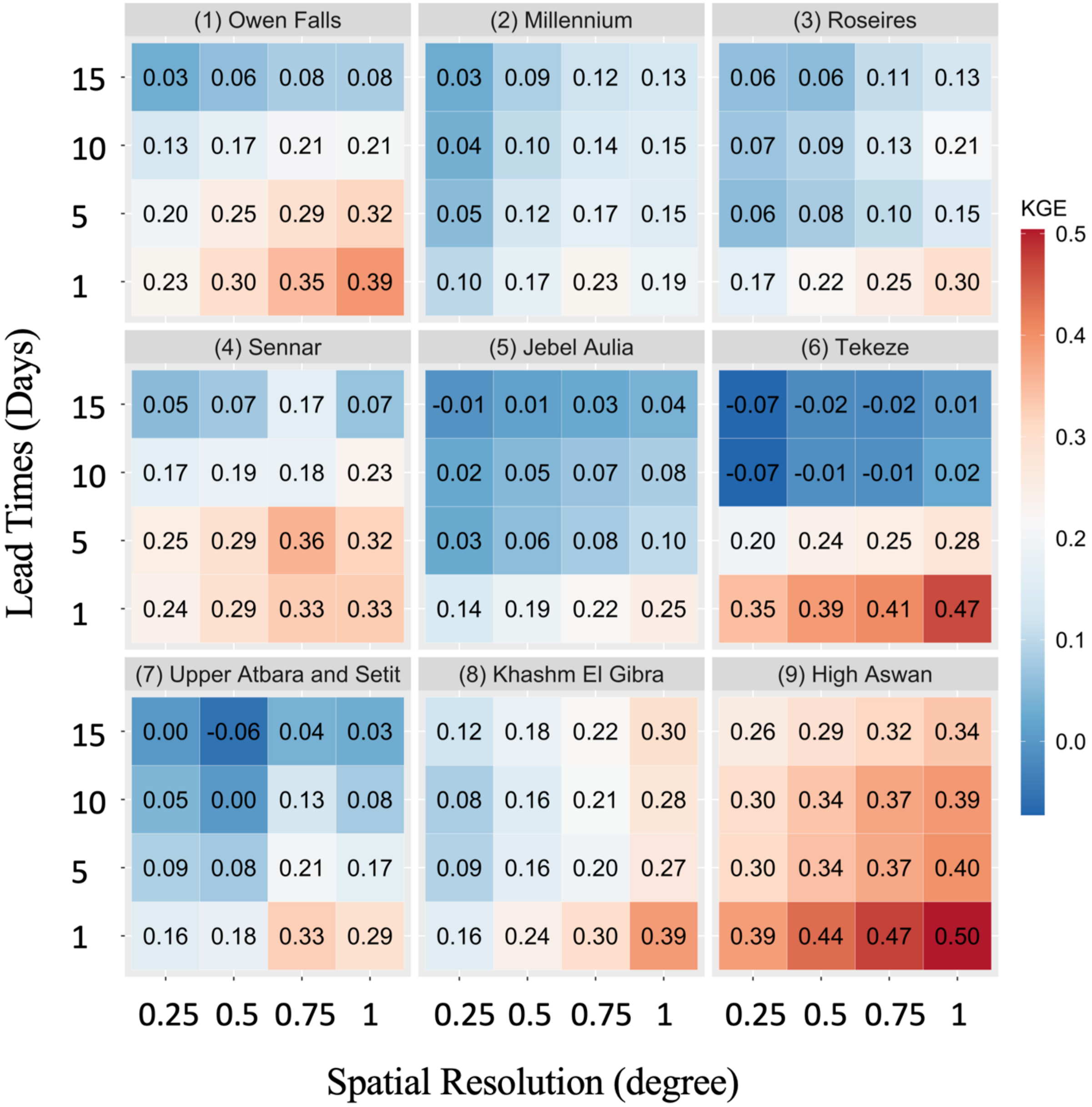


Figure 4.8. Kling-Gupta Efficiency (KGE) for daily precipitation of GFS compared to IMERG Final, for GFS lead times of 1-day, 5-day, 10-day, and 15-day, and at different spatial scales, for different Nile sub-basins.

With regard to the effect of lead time for daily forecasts, KGE decreases significantly as lead time increases. For instance, for the Owen Falls site, KGE decreases from 0.39 to 0.08 (1°, 1-day), and for Tekeze, it decreases from 0.47 to 0.01 (1°, 1-day). The correlation decreases as lead time increases. However, both bias ratio and variability ratio remain unaffected by the change in lead time. This suggests that longer lead time forecasts have similar bias and variability, but lower correlation compared to shorter lead time forecasts. The effect of spatial scale decreases with increasing lead time. For instance, for the Owen Falls site, while KGE increases by 0.16 from 0.23 (0.25°) to 0.39 (1°) for the 1-day lead forecast, it increases by only 0.05 from 0.03 (0.25°) to 0.08 (1°) for the 15-day lead forecast.

In Figure 4.9, we show the 5-day precipitation total, for three different lead periods, 1-5 day, 5-10 day, and 10-15 day. Overall, as the lead time increases, the KGE decreases, as expected. However, the KGE values for the 1-5 day lead and the 5-10 day lead are close to each other, while the KGE decreases considerably for the 10-15 day lead. Investigation of the KGE components reveals that the drop in KGE for the 10-15 day lead comes primarily from the drop in correlation between the forecasted and observed precipitation fields (see Figure 4.9c).

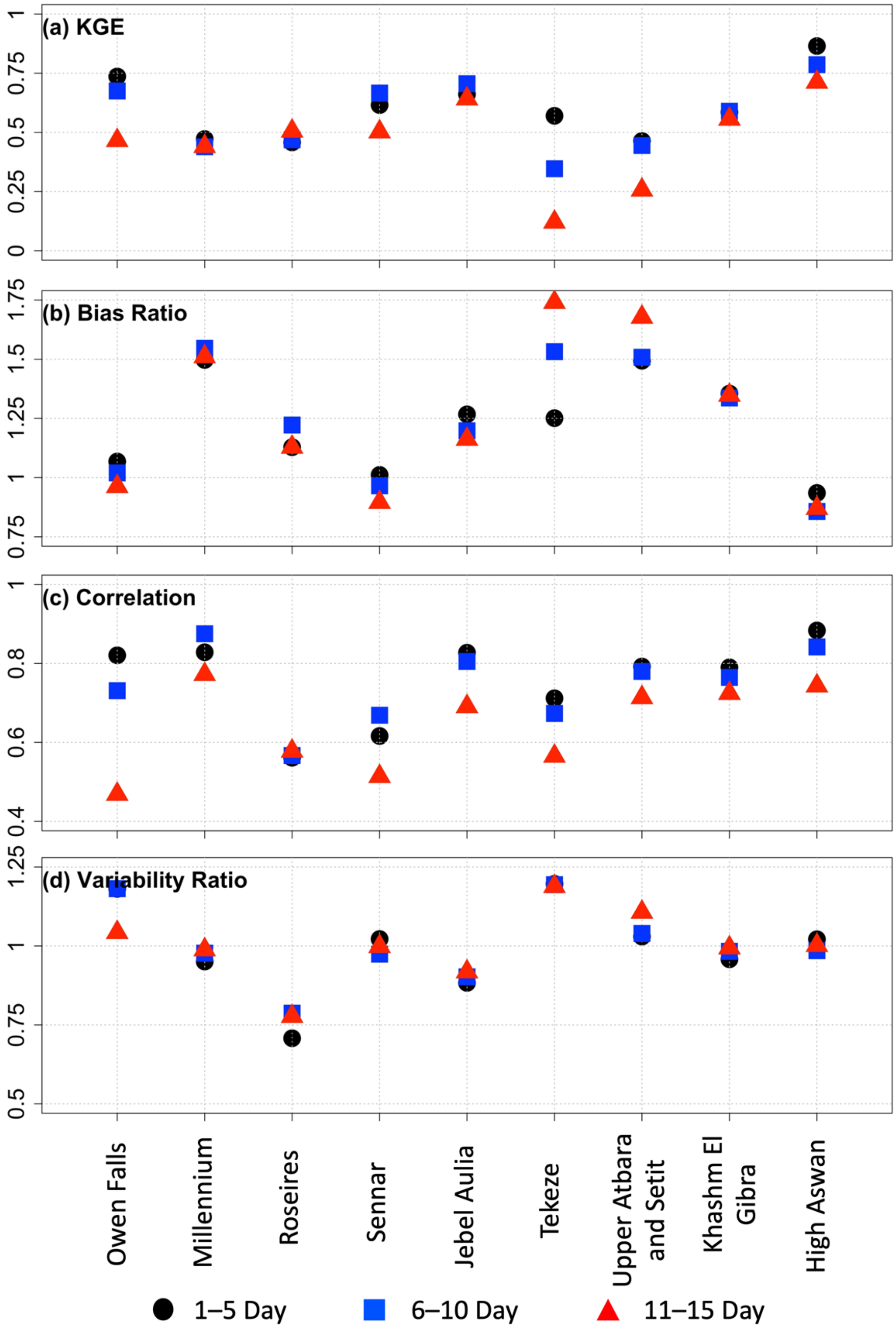


Figure 4.9. Kling-Gupta Efficiency (KGE) and its components for 5-day total rainfall forecast of GFS for three different lead time periods, 1-5 day, 5-10 day, and 10-15 day.

4.2.6 Effect of Temporal Aggregation Scale on Forecast Uncertainty

In order to assess the effect of temporal aggregation scale, we obtained the 1-day total, 5-day total, 10-day total, and 15-day total precipitation forecasts. These multi-day forecasts are constructed by combining multiple lead-time forecasts. For instance, the 5-day total forecast is obtained by adding the 1-day lead, 2-day lead, 3-day lead, 4-day lead, and 5-day lead daily forecasts. Figure 4.10 presents the KGE values for the multi-day-total basin-averaged forecasts as compared to IMERG Final reference.

In almost all cases (except Tekeze), the bigger the temporal aggregation scale is, the bigger the KGE. For instance, KGE increases from 0.19 (1°, 1-day-total) to 0.43 (1°, 15-day-total) for Millennium, and from 0.30 (1°, 1-day-total) to 0.74 for Roseires (1°, 15-day-total). The individual components of KGE are given in Supplemental Figures 4.4 – 4.6. As the temporal aggregation scale increases, correlation substantially increases in all cases. For example, correlation jumps from 0.44 (1°, 1-day-total) to 0.77 (1°, 15-day-total) for Millennium, and from 0.46 (1°, 1-day-total) to 0.85 (1°, 15-day-total) for Roseires. Similarly, the variability ratio improves with increasing temporal scale. For example, the variability ratio improves from 0.68 (1°, 1-day-total) to 0.99 (1°, 15-day-total) for Millennium, and from 0.57 (1°, 1-day-total) to 0.92 (1°, 15-day-total) for Roseires. On the other hand, the bias ratio remains mostly unaffected by temporal scale, except for Tekeze sub-basin, where the bias ratio worsens from 1.28 (1°, 1-day-total) to 1.55 (1°, 1-day-total).

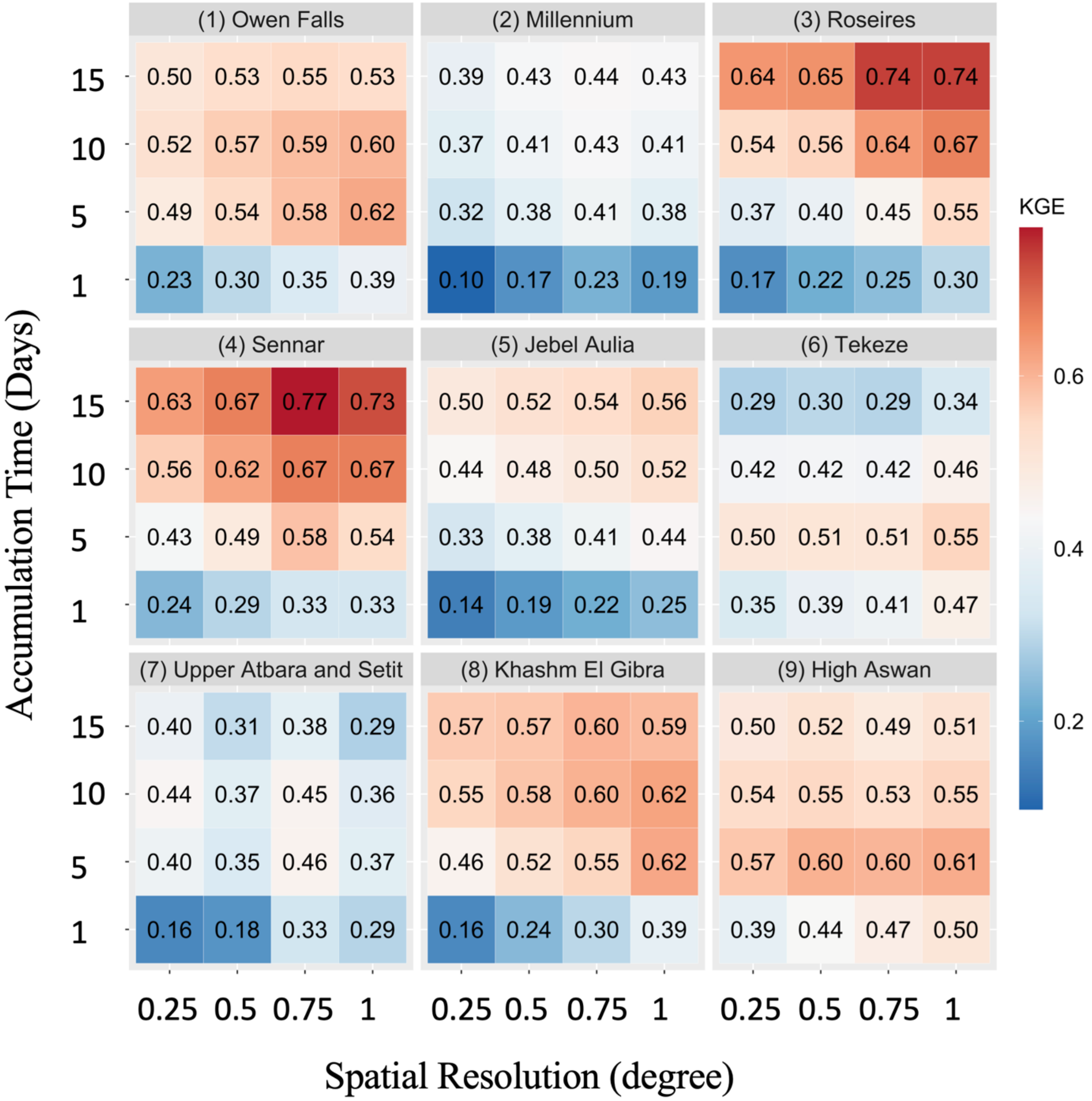


Figure 4.10. Kling-Gupta Efficiency (KGE) for GFS precipitation compared to IMERG Final, for accumulation time scales of 1-day, 5-day, 10-day, and 15-day, and at different spatial scales, for each Nile sub-basin.

In Figure 4.11, we show performance statistics for 15-day-total basin-averaged forecasts and estimates. In terms of KGE, two sub-basins show good estimates (High Aswan at 0.82 and Sennar at 0.80), three sub-basins show intermediate estimates (Roseires, Jebel Aulia, and Khashm El Gibrat), and four sub-basins show poor estimates (Owen Falls, Millennium, Tekeze, Upper Atbara and Setit). In terms of bias, five sub-basins show good estimates (Owen Falls, Roseries, Sennar, Jebel Aulia, and High Aswan), and four sub-basins show poor estimates (Millennium, Tekeze, Upper Atbara and Setit).

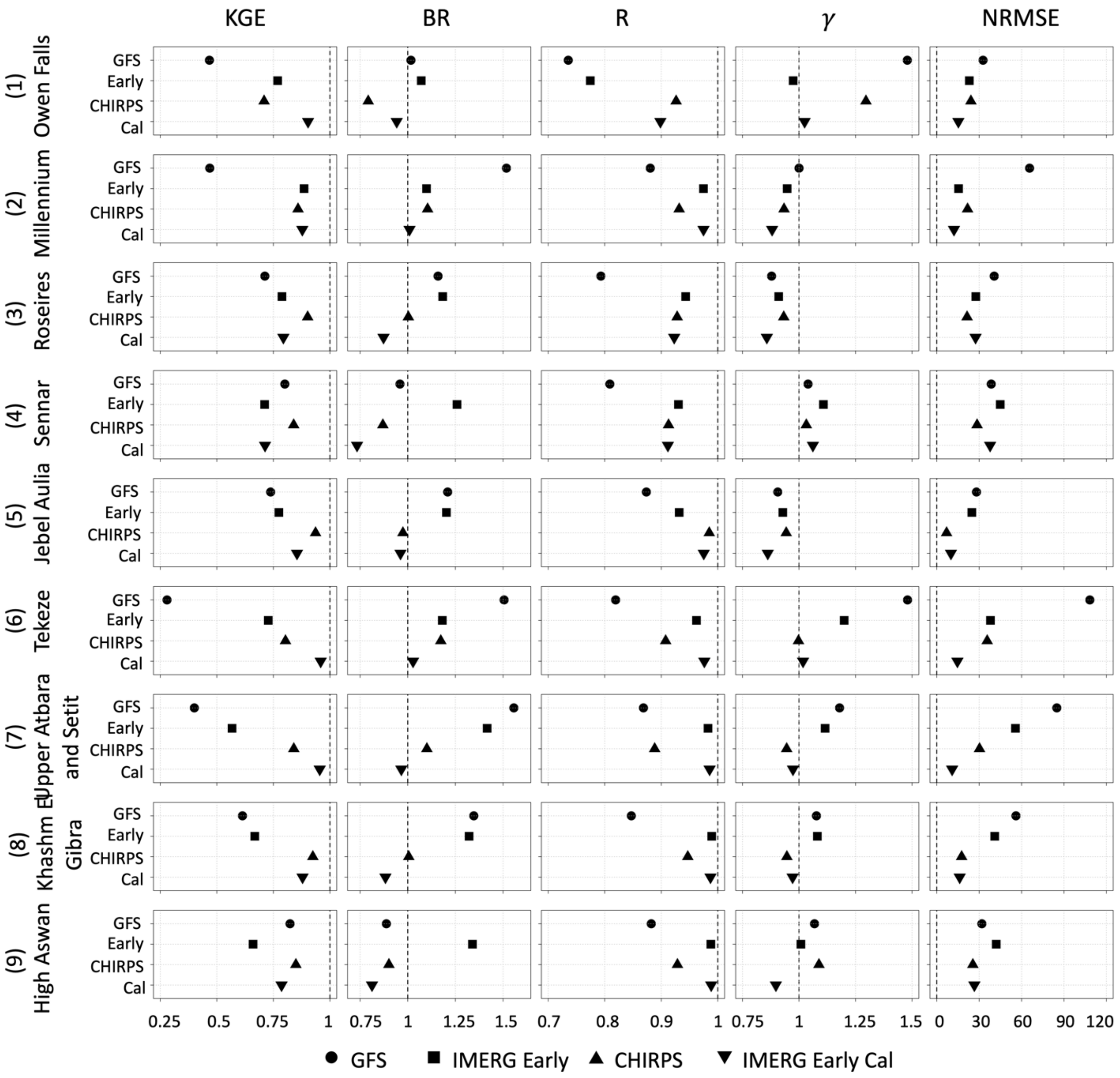


Figure 4.11. Summary of performance statistics (Kling-Gupta Efficiency KGE, Bias Ratio BR, correlation R, variability ratio γ, and root mean square error normalized by reference precipitation mean NRMSE), for sub-basin-averaged and 15-day accumulated rainfall derived from GFS forecasts and different satellite products, at different sub-basins.

4.2.7 How Well is the Spatial Pattern of Daily Precipitation Forecasted?

The spatial map of annual precipitation from June 2019 to June 2020 is shown in Figure 4.12. According to IMERG Final, the southern and eastern parts of the basin are the wettest, while the northernmost part is the driest. GFS captures well the overall spatial pattern of annual precipitation with a correlation coefficient of 0.92 (with IMERG Final) but tends to overestimate in the wettest parts of the basin (Millennium, Owen Falls, Upper Atbara & Setit, ad Khashm El Gibrat). IMERG Early also captures well the spatial pattern with a correlation coefficient of 0.98 but tends to overestimate precipitation in the central part (Jebel Aulia) where the precipitation is moderate. CHIRPS also shows high correlation (0.95) with IMERG Final, but tends to underestimate precipitation in the southern part (Owen Falls).

The time series of correlation between the spatial (0.25° grids across the Nile basin) distribution of IMERG Final reference and the rest of precipitation products is shown for daily precipitation (Figure. 4.13a) and 15-day-total precipitation (Figure. 4.13b). At the daily time scale, the correlation between GFS and IMERG Final is mostly around 0.40 ± 0.14. For 15-day accumulation, the correlation between GFS and IMERG Final jumps to 0.76 ± 0.09. For IMERG Early, the correlation with IMERG Final is close to one, for both daily and 15-day accumulation. Bias correction of IMERG Early would not increase the correlation any further. As far as CHIRPS is considered, the correlation with IMERG Final is around 0.58 ± 0.12 for daily time scale and 0.86 ± 0.04 for 15-day accumulation time scale.

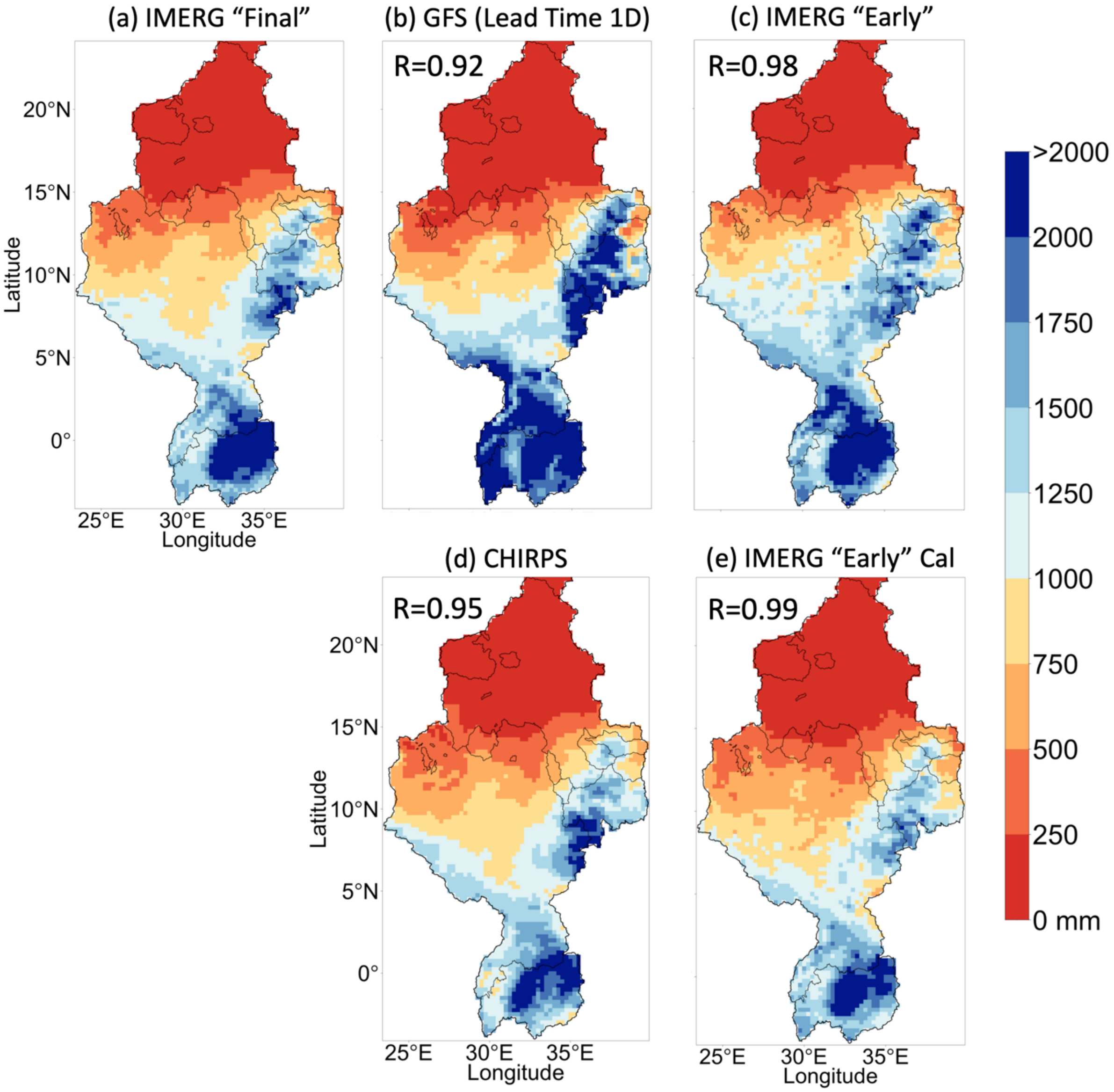


Figure 4.12. Spatial map of annual precipitation products obtained from (a) IMERG Final, (b) 1-day lead GFS forecast, (c) IMERG Early, (d) CHIRPS, and (e) Climatologically-bias-corrected IMERG Early (IMERG Early Cal), across the Nile basin.

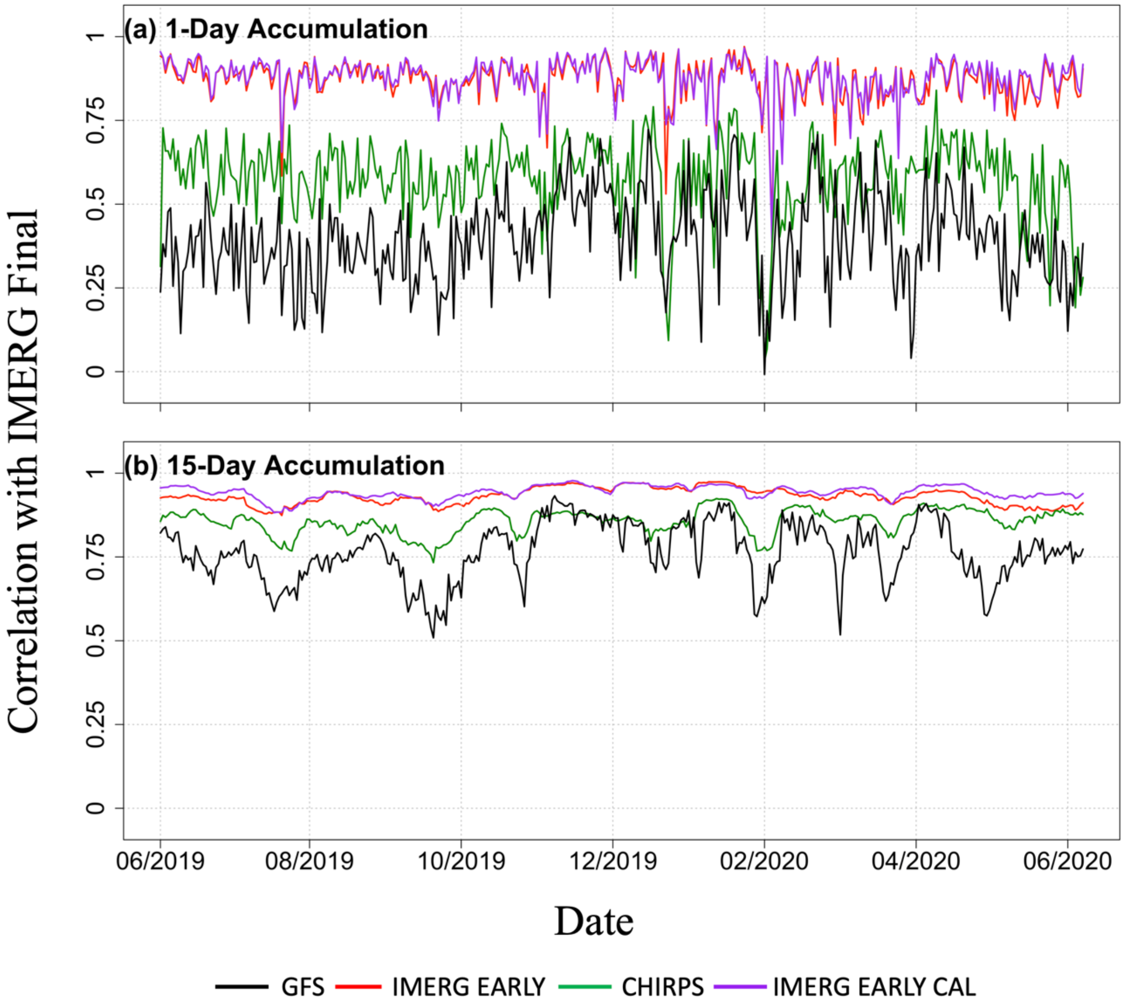


Figure 4.13. Spatial correlation between the daily basin-averaged rainfall products (1-day lead GFS, IMERG Early, CHIRPS, climatologically-bias-corrected IMERG Early) and IMERG Final, for the Nile basin.

4.3 Summary and Conclusions

In this study, we assessed the performance of medium-range (1-day to 15-day lead time) forecasts available from the Global Forecast System (GFS), for the sub-basins of the Nile River Basin. The motivation of this work stemmed from the need to evaluate the skills of forecasts so they can be used in confidence in applications, such as reservoir operation and management, and to identify strengths and weaknesses of the forecast quality in order to provide insights to algorithm developers and product users. The Nile River sub-basins are defined here on the basis of major reservoir dam locations, namely, Owen Falls, Millennium, Roseires, Sennar, Jebel Aulia, Tekeze, Upper Atbara and Setit, Khashm El Gibra, and High Aswan.

The overall performance of the 1-day lead-time basin-averaged GFS forecast, as measured through KGE, is poor ($0 < KGE < 0.5$) for the majority of the sub-basins (7 sub-basins), namely Millennium, Roseires, Sennar, Jebel Aulia, Tekeze, Upper Atbara and Setit, and Khashm El Gibra. The KGE is better (KGE of 0.69 and 0.74) for two sub-basins, namely Owen Falls, and High Aswan. The KGE breakdown indicates that the nature of forecast quality is not the same for all sub-basins. At the daily timescale, the 1-day lead basin-averaged GFS forecast has large overestimation bias for three watersheds (Millennium at 47%, Upper Atbara and Setit at 50%, and Khashm El Gibra at 42%). The correlation between the 1-day lead basin-averaged GFS and IMERG Final is poor (correlation smaller than 0.60) for five sub-basins, particularly for Sennar (0.36) and Roseires (0.45), but was good for the remaining four sub-basins (High Aswan at 0.75, Jebel Aulia at 0.73, Owen Falls at 0.72, and Millennium at 0.68). The variability ratio of the GFS is close to the reference precipitation for most sub-basins, but is too low (meaning smoother precipitation variability) for Roseires (variability ratio of 0.56) and Jebel Aulia (variability ratio of 0.66).

Overall, the daily and 1-day lead GFS forecast uncertainty varies depending on a combination of factors, mainly, watershed elevation and area, climatological regime, rain rate, forecast lead time and accumulation timescale. The wet climatological regimes are located in the northeastern parts of the Nile (namely, Millennium, Upper Atbara & Setit, and Khashm El Gibra watersheds) and the southern part of the Nile (Owen Falls watershed). Results show that the GFS forecasts have low bias in the drier parts of the basin, and low or high bias in the wet parts of the basin depending on the hemisphere. The watersheds located in a wet climatic regime in the northern hemisphere show large overestimation bias, namely, the Millennium watershed (i.e. the watershed for the newly-built Great Ethiopian Renaissance Dam (GERD)), which is a highly mountainous area, and the Upper Atbara & Setit as well as the Khashm El Gibra watersheds, which are adjacent medium-elevation watersheds. Whereas, the watershed located in a wet climatic regime in the southern hemisphere, i.e., Owen Falls watershed, which is a medium-elevation watershed, shows low bias. In terms of capturing the temporal dynamics of observed watershed-averaged precipitation, larger watersheds show better skills (as measured through correlation) than relatively smaller watersheds. As the lead time increases from 1-day to 15-day, the forecast accuracy decreases. In particular, the forecast accuracy decreases considerably for lead times exceeding 10 days. Averaging the forecasts at coarser spatial and temporal scales leads to increased forecast accuracy. We especially point out that the improvement in accuracy at bigger temporal averaging periods comes despite the fact that the forecast accuracy decreases with increasing lead time.

The study also looked into the potential of the near-real-time product IMERG Early for use to calibrate and further improve the accuracy of GFS forecasts, by simply comparing the performance of IMERG Early with the performance of GFS forecasts. For Millennium watershed (one of the three watersheds showing large overestimation bias in GFS forecasts), IMERG Early has better bias than the GFS forecast. Whereas, for the Upper Atbara & Setit and Khashm El Gibrat watersheds, IMERG Early has similar level of bias as with GFS forecasts. The relative performance of IMERG Early could be partly explained by the number of rain gauges used in the IMERG Final: 16 rain gauges for the Millennium watershed, but only one rain gauge each for the Upper Atbara & Setit and Khashm El Gibrat watersheds. In addition, in the case of the driest watershed in the basin (i.e. the High Aswan), IMERG Early (despite having 16 rain gauges) shows large overestimation bias, which is worse than the GFS forecasts. The IMERG Early and GFS forecast have similar bias in the remaining watersheds.

Climatological bias correction of IMERG Early (obtained by calculating the bias between the long-term average IMERG Final and IMERG Early estimates at each pixel, and adjusting the IMERG Early estimates during the study period based on the bias) has improved the bias in IMERG Early, particularly it has reduced the large bias for Upper Atbara & Setit (from +40% to -5%) and for Khashm El Gibrat (from +31% to -12%). However, the climatological-bias correction has not improved the bias in IMERG Early for some basins. We recommend developing methods to improve the accuracy of GFS forecasts, including appropriate post-processing techniques that use IMERG-Early and IMERG Final products. Post-processing techniques that could be considered include multi-resolution bias correction through wavelet analysis or empirical mode decomposition method, and Artificial-based methods such as Feed Forward Neural Network, Support Vector Machine, and Adaptive Neural Fuzzy Inference System.

CHAPTER 5

Evaluation of Medium-Range Forecasts (GFS) over Niger

This chapter has been accepted in the Hydrology and Earth System Sciences. Used with permission.

Yue, H., Gebremichael, M. and Nourani, V., 2021. Performance of the Global Forecast System's Medium-Range Precipitation Forecasts in the Niger River Basin. Hydrology and Earth System Sciences Discussions, pp.1-31, accepted.

5.1 Study Region

The Niger River is the principal river of West Africa, and is shared among nine riparian countries (Figure. 5.1): Benin, Burkina Faso, Cameroon, Chad, Guinea, Ivory Coast, Mali, Niger and Nigeria. The basin is facing multiple pressures from increasing population, water abstraction for irrigation, and risk of extreme hydrological events due to climate change (Sylla et al. 2018). A number of hydropower dams exist in the region, and additional dam projects are envisaged in order to alleviate chronic power shortages in the countries of the Niger basin.

The Niger river, with a drainage basin of 2,117,700 Km2, is the third longest river in Africa. The source of the main river is in the Guinea Highlands, and runs through Mali, Niger, on the border with Benin and then through Nigeria, discharging through a massive delta, known as the Niger Delta (the world's third largest wetland), into the Atlantic Ocean. The rainfall regimes in the region follow the seasonal migration of the Inter-Tropical Convergence Zone

(ITCZ), which brings rainfall primarily in the summer season (Animashaun et al. 2020; Sorí et al. 2017). Climatologically, the Niger basin lies in three latitudinal sub-regions (Akinsanola et al. 2015, 2017): (1) the Guinea coast (latitude 4°–8°N), which borders the tropical Atlantic Ocean in the south; (2) the Savannah (latitude 8°–12°N), an intermediate sub-region; and (3) the Sahel (latitude > 12°N) to the north. The Guinea coast experiences a bimodal rainfall regime that is centered in the summer monsoon period of June–September, with August being the period of a short dry season, while the Savannah and Sahel sub-regions experience a unimodal rainfall regime, with maximum rainfall occurring in August (Akinsanola and Zhou 2018). The ranges of annual rainfall amounts are: 400–600 mm in the Sahel, 900–1200 mm in the Savannah; and 1500–2000 mm in the Guinea coast (Akinsanola et al. 2017).

The Niger basin is home to eight major reservoir dams (see Table 5.1 and Figure 5.1): (1) Selingue Dam in Mali: a primarily hydropower dam, (2) Markala Dam in Mali: a primarily irrigation dam, serving about 75,000 ha of farmland, (3) Goronyo Dam in Nigeria: a multi-purpose dam for flood control, provision of downstream water supply and the release of water for irrigation in the dry season, (4) Bakolori Dam in Nigeria: a primarily irrigation dam with a command area of about 23,000 ha, (5) Kainji Dam in Nigeria: the largest Dam on the Niger supplying power for most towns in Nigeria, (6) Jebba Dam in Nigeria: a primarily hydropower dam, (7) Dadin Kowa Dam: a multi-purpose dam for water supply, electricity and irrigation, (8) Lagdo Dam in Cameroon: a multi-purpose dam providing electricity to the northern part of the country and supplying irrigation water for 15,000 hectares of cropland. The watersheds of the dams are primarily either in the Savanna (Selingue, Markala, Jebba, Dadin Kowa, an Lagdo), or in the Sahel (Goronyo, Kainji), or partly in both (Bakolori). The watershed sizes vary over a large range, from 4,887 Km2 (Bakolori Dam) to 1,464,092 Km2 (Kainji Dam). The average elevations of the watersheds are close to each other at 500 ± 50 m.a.s.l.

In order to make the results of this study meaningful to reservoir managers, the Niger basin was divided into watersheds according to the locations of the dam reservoirs (see Fig. 1). Then the sub-basin of each dam was defined as the drainage between the dam itself and the upstream dam. For example, the drainage basin of the Markala Dam does not include the drainage basin of the Selingue Dam.

Table 5.1. Selected dams and their watershed characteristics

Dam	Country	Operational since *	Capacity (million m^3)*	Power (MW) *	Primary Purpose*			Area of Drainage Basin (km^2)**	Elevation of Drainage Basin (m)**
					Irrigation and Water Supply	Flood Control	Hydroelectricity		
Selingue	Mali	1982	2170	44			x	32685	473
Markala	Mali	1947	175		x			102882	442
Goronyo	Nigeria	1983	942		x	x		31547	446
Bakolori	Nigeria	1978	450		x			4887	519
Kainji	Nigeria	1968	15000	960			x	1464092	406
Jebba	Nigeria	1984	3600	540			x	40268	308
Dadin Kowa	Nigeria	1988	2855	35	x		x	32936	535
Lagdo	Cameroon	1983	7800	72		x	x	31352	452

* information obtained from the Global Reservoir and Dam Database (Lehner et al. 2011) and Food and Agriculture Organization of the United Nations (FAO)'s Global Information System on Water and Agriculture (AQUASTAT).
** Calculated from HydroSEHDS (Lehner et al. 2008).

5.2 Results and Discussion

5.2.1 Annual Spatial Variability and Seasonal Characteristics

The spatial map of annual (15 June 2019 – 15 June 2020) rainfall from the various rainfall products is given in Figure 5.2. According to the reference rainfall product (i.e. IMERG Final), the Niger basin experiences average annual rainfall of 700 mm. The spatial rainfall distribution shows north-to-south increasing gradient, with the Sahel region (> 12°N) receiving on average 346 mm per year, the Savanna region (8°N – 12°N) receiving on average 1,206 mm per year, and the Guinea region (4°N – 8°N) receiving on average 1,620 mm per year. The spatial structures (climatology and north-south gradient in rainfall) of GFS, IMERG and CHIRPS rainfall fields are quite similar to those of IMERG Final. However, the 1-day GFS tends to overestimate in the wet Guinea region of the basin, whereas both IMERG Early and CHIRPS give values that are very close to IMERG Final.

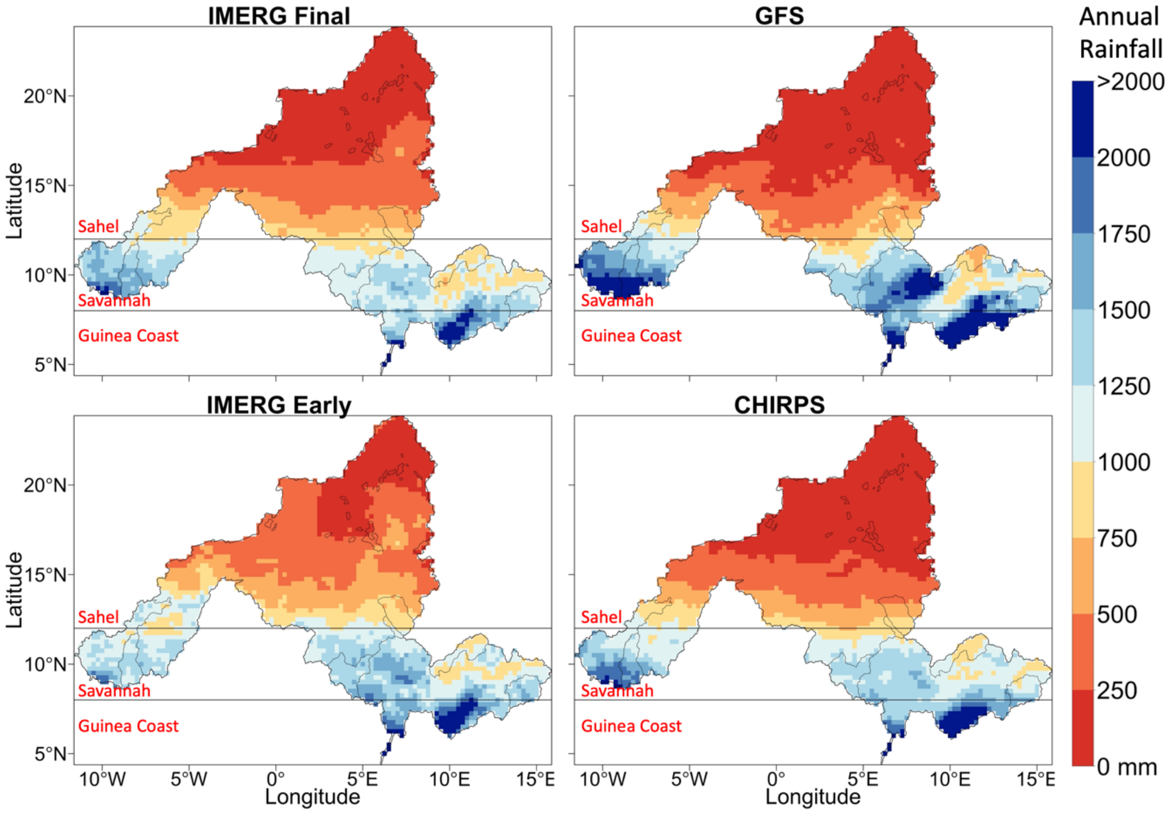


Figure 5.1. Spatial map of annual rainfall (in mm), for the period 15 June 2019 to 15 June 2020, derived from (a) IMERG Final, (b) GFS (1-day lead time), (c) IMERG Early, and (d) CHIRPS.

Figure 5.3 shows the seasonal rainfall pattern for each climatological region. According to the reference IMERG Final, as one goes from north to south, the rainy season expands from 3 months (June – September) in the Sahel to 6 months (March – November) in the Savanna and Guinea regions. The peak rainfall also shows north-south gradient, with peak rainfall of 130 mm in the Sahel, to 269 mm in the Savanna, and 350 mm in the Guinea. The rainfall pattern is unimodal with a peak rainfall value in August for both Sahel and Savanna, but becomes bimodal with one peak in May and the other in September for Guinea.

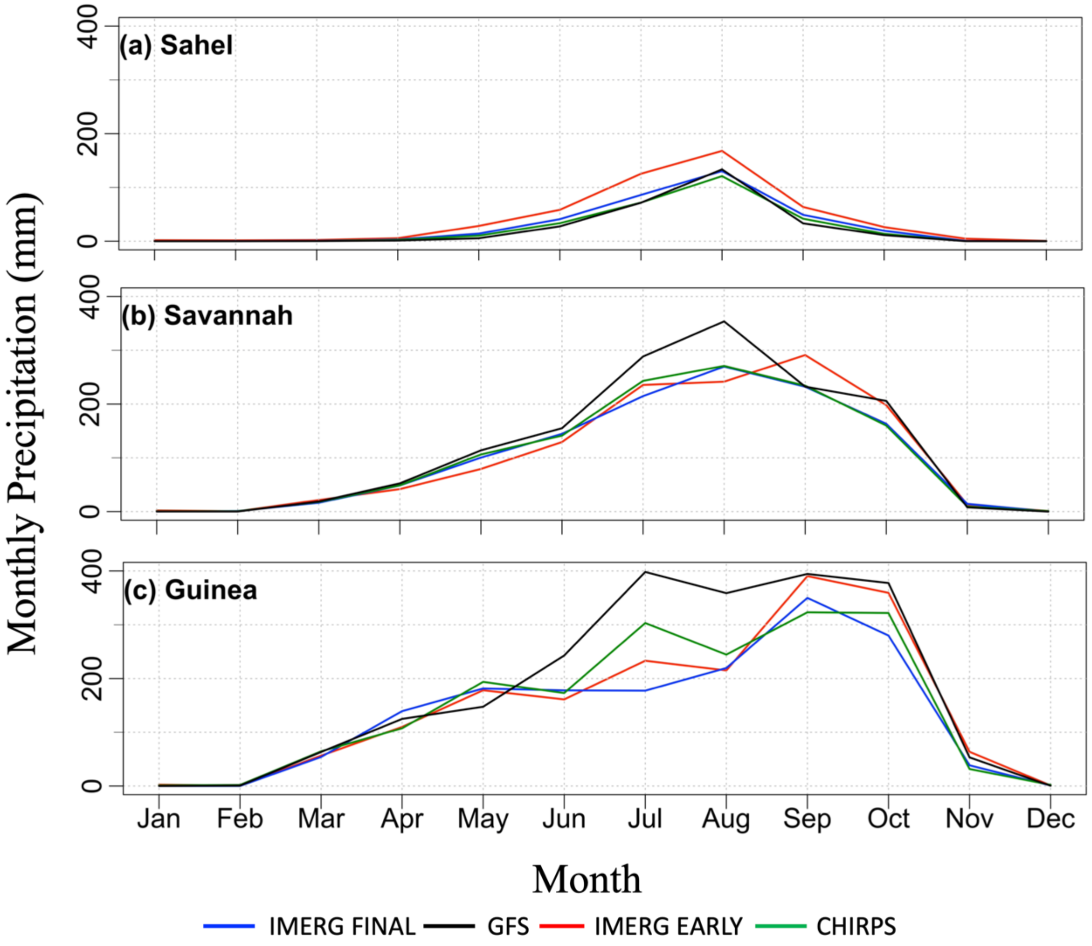


Figure 5.2. Monthly precipitation regime for the three climatological zones of the Niger river Basin: (a) Sahel, (b) Savanna, and (c) Guinea. Analyses are based on rainfall fields derived from IMERG Final, 1-day-lead GFS. IMERG Early, and CHIRPS. The time period covers from 15 June 2019 to 15 June 2020.

When validated against IMERG Final, the performance of GFS in capturing the seasonal rainfall characteristics deteriorates as one goes from north to south. GFS captures both the seasonal rainfall pattern and rainfall peak in the Sahel, and captures the seasonal rainfall pattern but tends to moderately overestimate the peak in the Savannah, while it has large overestimation (almost twice as much as the reference) in the Guinea particularly during summer. As far as the other satellite products are concerned, IMERG Early tends to slightly overestimate in the Sahel across all rainy months, but performs relatively well in the Savannah and Guinea regions. CHIRPS is very close to IMERG Final in all regions and months, with the exception of modest overestimation of the July rainfall in Guinea.

5.2.2 How Well do GFS Forecasts Capture Annual Rainfall?

Here, we aggregate the 1-day lead GFS forecasts to annual time scale and compare the results against corresponding annual precipitation estimates from IMERG Final. Figure 5.4 presents the watershed-averaged annual rainfall for each dam watershed. According to IMERG Final, the annual rainfall varies from 434 mm (in Kainji) to 1,481 mm (in Selingue). Watersheds 1 (Selingue) and 2 (Markala), located in the western part of the Savannah, receive the largest amount of rainfall, i.e., 1481 mm and 1406 mm, respectively. Watershed 3 (Markala), located in the eastern part of the Sahel, receives 741 mm of annual rainfall. Watershed 4 (Bakolori), characterized by the smallest watershed area compared to the rest of the watersheds, lies partly in the Sahel and partly in the Savannah region and receives 921 mm of annual rainfall. Watershed 5 (Kainji), characterized by the largest watershed area of all, lies mostly in the Sahel region and receives the lowest amount of annual rainfall (434 mm). Watersheds 6 (Jebba), 7 (Dadin Kowa), and 8 (Lagdo), located in the Savannah, receive annual rainfall amounts of 1190 mm, 941 mm, and 1295 mm, respectively.

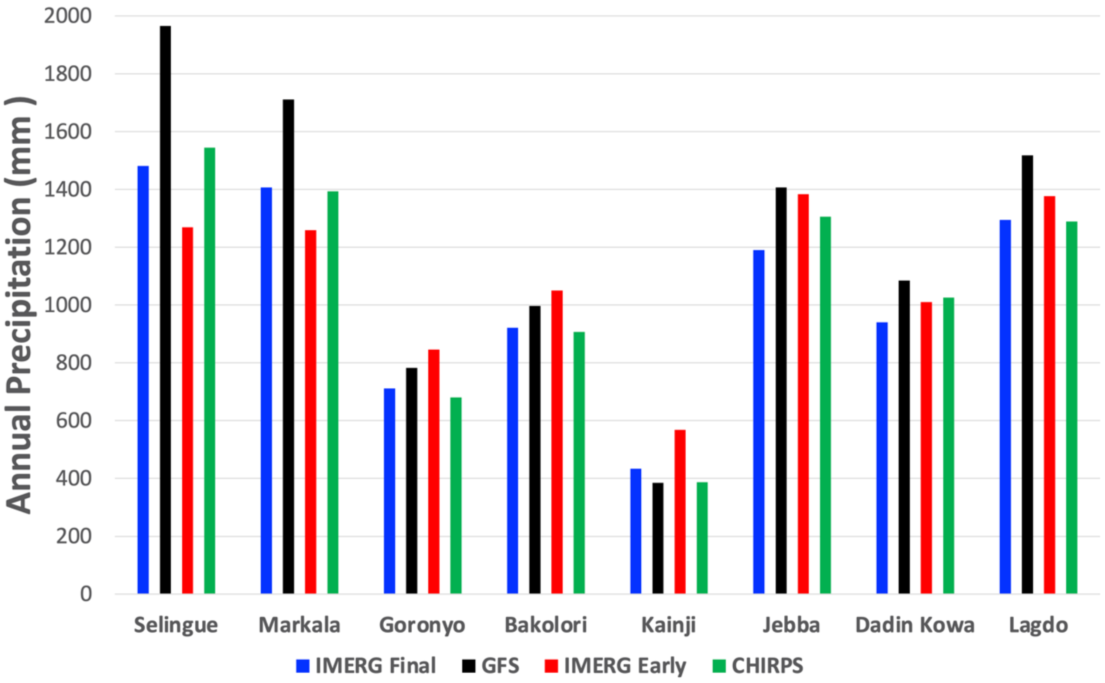


Figure 5.3. Sub-basin averaged annual precipitation (mm) for the period, 15 June 2019 to 15 June 2020, for each of the Niger's sub-basin, derived from the 1-day lead GFS forecast and different satellite precipitation products.

Validated against IMERG Final, the GFS tends to overestimate rainfall in all watersheds located in the Savannah (or watersheds that receive relatively large rainfall amounts), with an overestimation varying in the range 8% to 33%, with larger bias for watersheds receiving higher rainfall amount. For watersheds in the Sahel (watersheds receiving low rainfall amount), GFS gives less bias (-11% for the driest Kainji watershed and +10% for Bakolori).

In contrast, IMERG Early tends to underestimate rainfall in all watersheds located in the Savannah (with larger negative bias in watersheds with large rainfall amount) but tends to overestimate in all watersheds located in the Sahel (with very large overestimation bias for the driest watershed) Therefore, GFS and IMERG Early have different bias characteristics: whereas GFS outperforms IMERG Early in the Sahelian climate where well-organized convective systems dominate the monsoon, IMERG Early outperforms GFS in the Savannah and Guinea climate which are characterized by short-lasting and localized systems and wet land surface conditions. CHIRPS estimates are reasonably close to IMERG Final, indicating that the choice of reference product between CHIRPS and IMERG Final would not substantially affect the findings on the accuracy of GFS forecasts.

5.2.3 How well is the Time Series of Daily Precipitation Forecasted?

Figures 5.5 and 5.6 present the time series of watershed-averaged daily rainfall, for the wet period June – October. According to IMERG Final, the temporal variability (as measured through coefficient of variation or CV) varies from 1.22 to 2.60. Validated against IMERG Final, the GFS tends to underestimate the temporal variability and particularly underestimate large spikes in rainfall, at almost all sites except at Kainji. The GFS's relatively better performance for Kainji could be attributed to the watershed's large area that results in relatively smooth temporal variability. Both IMERG Early and CHIRPS provide CV values that are very close to IMERG Final.

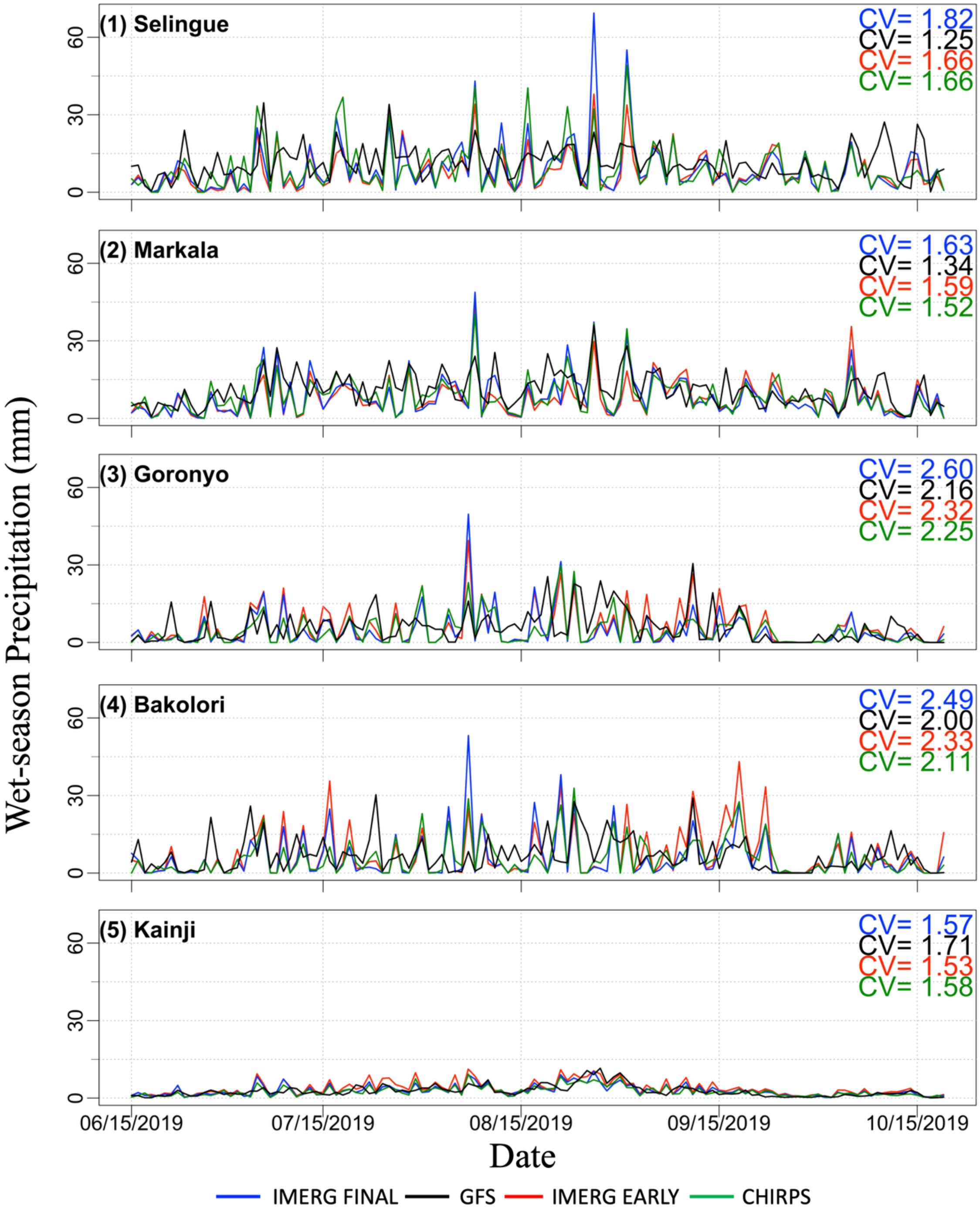


Figure 5.4. Time series of sub-basin averaged precipitation total (mm) for the wet period (June – September 2019 for all sub-basins, derived from various precipitation products, for five sub-basins. The Figure also shows the coefficient of variation (CV) as a measure of temporal variation.

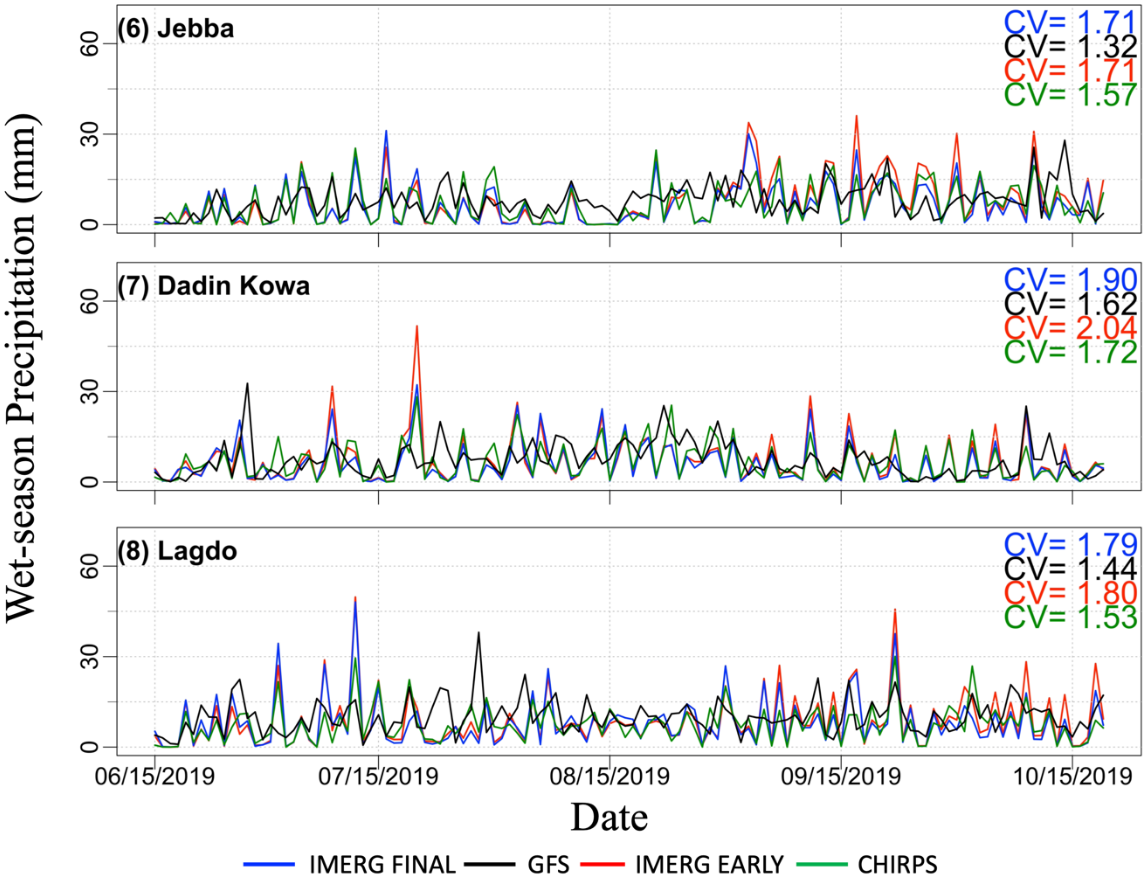


Figure 5.5. Same as in Figure 5.5 but for the remaining three watersheds.

Figure 5.7. displays the performance statistics of watershed-averaged daily rainfall (validated against IMERG Final) in terms of Kling-Gupta Efficiency (KGE), Bias Ratio (BR), correlation (R), variability ratio (γ), and root mean square error normalized by reference precipitation mean (NRMSE). First, the performance results for the 1-day lead GFS are considered. The KGE scores are poor ($0.3 < KGE < 0.5$) for half of the watersheds considered (Selingue, Goronyo, Bakolori, and Lagdo) and intermediate ($0.5 < KGE < 0.75$) for the remaining half watersheds (Markala, Kainji, Jebba, and Dadin Kowa). The breakdown of the KGE scores (BR, R, and γ) reveals the key factors contributing to the KGE estimates. The GFS

tends to overestimate daily precipitation for most sub-basins, as BR is higher than one, except for Kainji. The overestimation is particularly high for Selingue and Markala, where BR is 1.33 and 1.22, respectively. The correlation coefficient between GFS and IMERG Final is mostly low (R < 0.60), and is particularly lower for Bakolori (R=0.36) and Goronyo (R=0.43). The variability ratio of GFS is mostly between 0.69 to 0.83 (except for Kainji, where γ is 1.09), indicating that the GFS tends to give lower temporal variability of rainfall. The NRMSE is very high, ranging from 100% to 266%, and is particularly high for Goronyo (266%) and Bakolori (264%), which are relatively small-sized watersheds.

Next, the performance of IMERG Early was examined with respect to IMERG Final, mainly to assess if it is possible to use the near-real-time IMERG Early product to calibrate and improve the accuracy of GFS forecasts. The IMERG Early performs much better with KGE values higher than 0.75 (except for Kainji where KGE is 0.69), correlation higher than 0.90, and variability ratio close to the optimum value. The high performance of IMERG Early is due to its similarity with the IMERG Final product, as the main difference between the two products is that IMERG Early, unlike IMERG Final, does not use monthly rain gauge observations for bias correction. Such monthly bias correction techniques would not alter the pattern and variability of IMERG Early compared to IMERG Final. Therefore, the performance of IMERG Early should be evaluated using bias statistics, the other statistics (correlation and variability ratio) are presented for completeness. IMERG Early overestimates rainfall in most watersheds in the range 11% (Lagdo) to 28% (Kainji) except for two watersheds, where it slightly underestimates by 14% (Selingue) and 11% (Markala). Comparison of the performance of GFS and IMERG Early indicates that both products have different bias characteristics. In some watersheds (e.g., Kainji), GFS outperforms IMERG Early in terms of bias, whereas in other watersheds (e.g., Markala), IMERG Early outperforms GFS.

CHIRPS was also compared with IMERG Final to assess how the use of different reference products may affect the finding about the performance of GFS forecasts. The KGE scores of CHIRPS are higher than 0.75 in all cases, indicating that CHIRPS and IMERG Final have comparable KGE performance. Therefore, the performance of GFS is expected to be about the same even if the reference product used this in this study (IMERG Final) changes to CHIRPS.

5.2.4 Dependence of Forecast Performance on Precipitation Rate

Figure 5.8 presents the scatterplot of 1-day lead GFS forecasts and IMERG Final at daily and watershed-average scales. The performance of GFS varies between watersheds. In the Markala and Kainji watersheds, GFS forecasts agree well with IMERG Final at almost all rain rates. In the Selingue watershed, GFS agrees well with IMERG Final for rain rates under 30 mm/day, but GFS substantially underestimates all rain rates above 30 mm/day. In the remaining five watersheds, GFS has poor performance, replete with large scatter, high false alarm, and large underestimation bias of heavy rain rates.

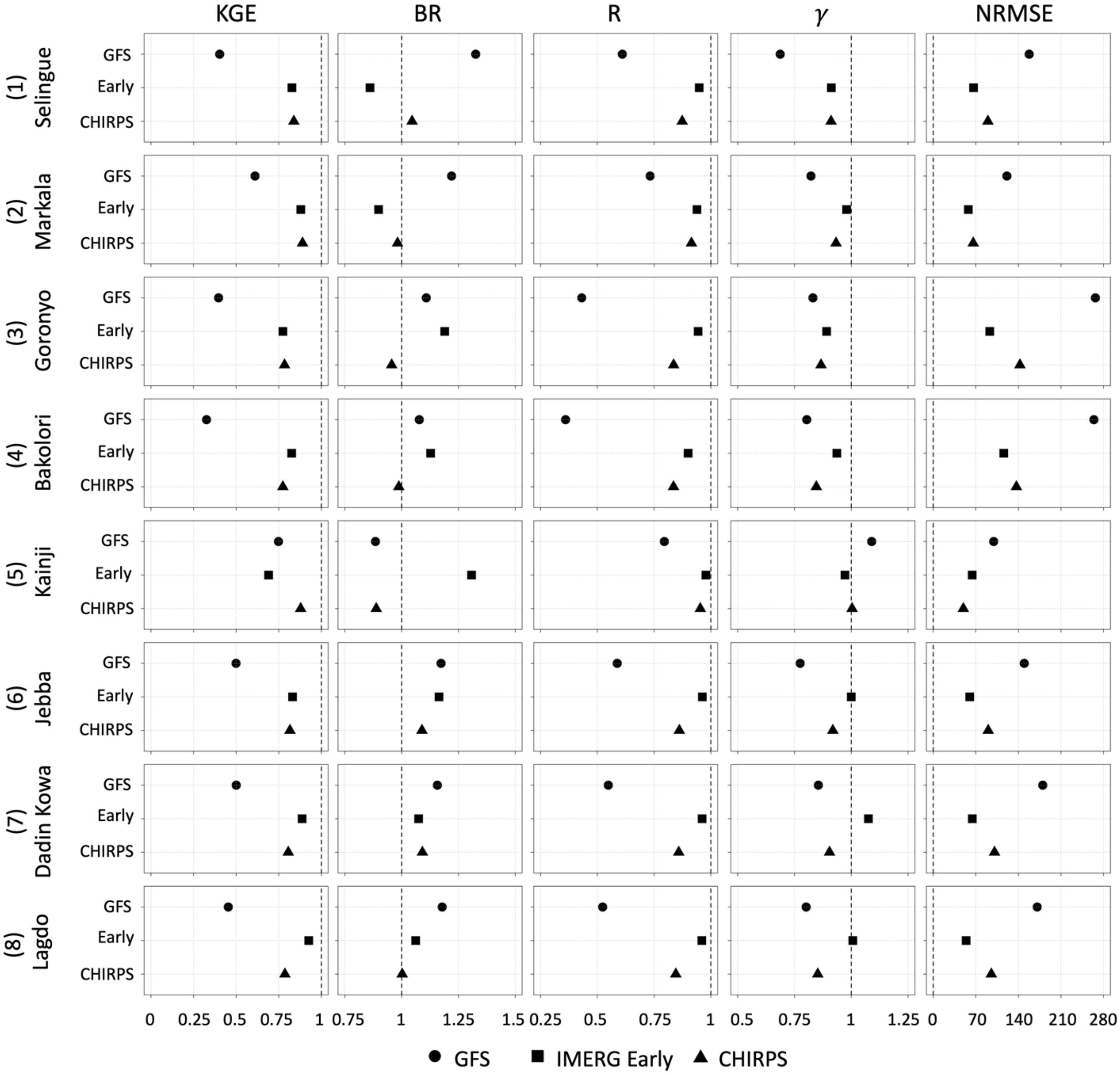


Figure 5.6. Summary of performance statistics (Kling-Gupta Efficiency KGE, Bias Ratio BR, correlation R, variability ratio γ, and root mean square error normalized by reference rainfall [%], for the 1-day lead time GFS forecasts and other satellite products. The time period considered was June 15, 2019 – June 15, 2020.

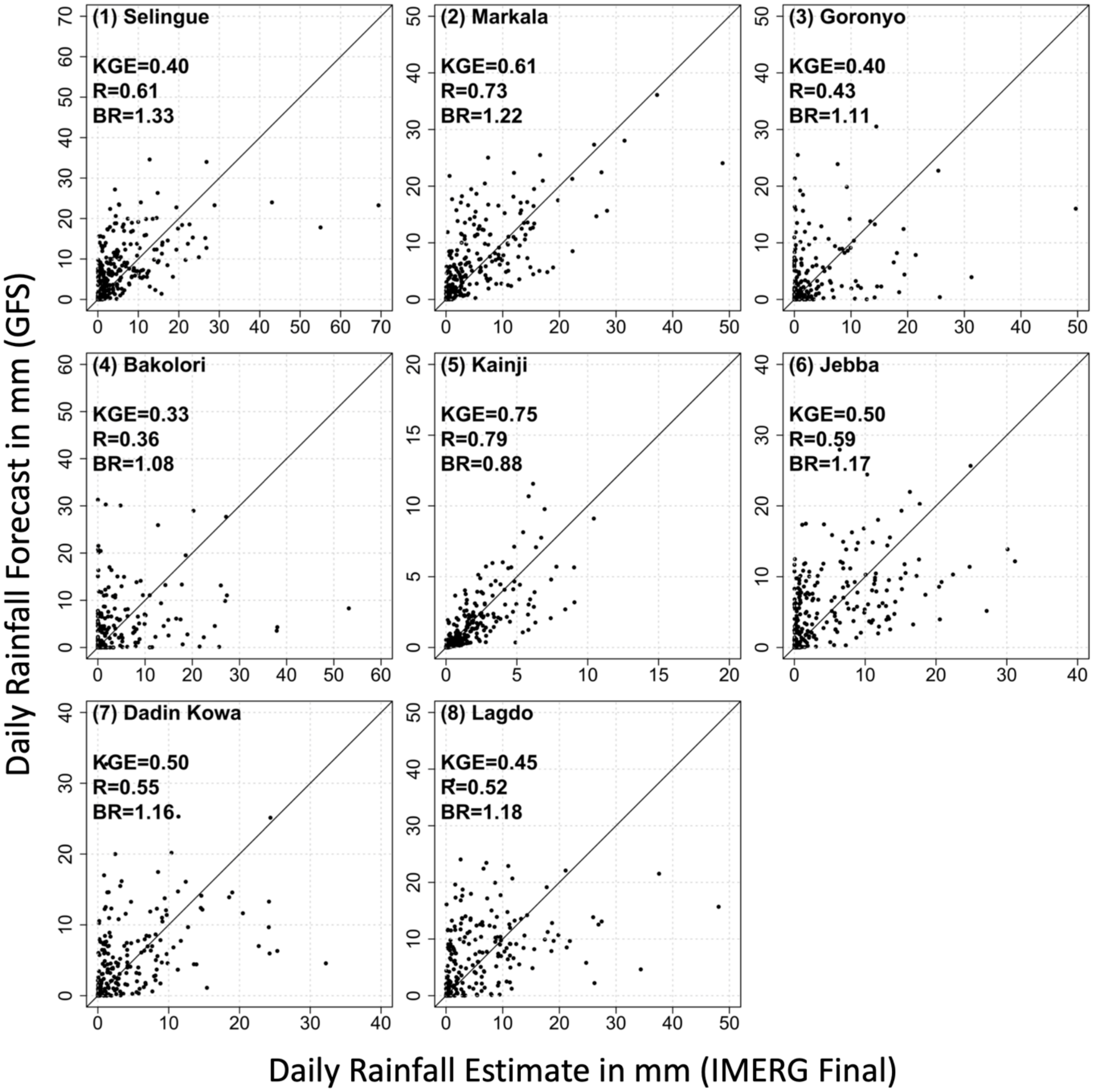
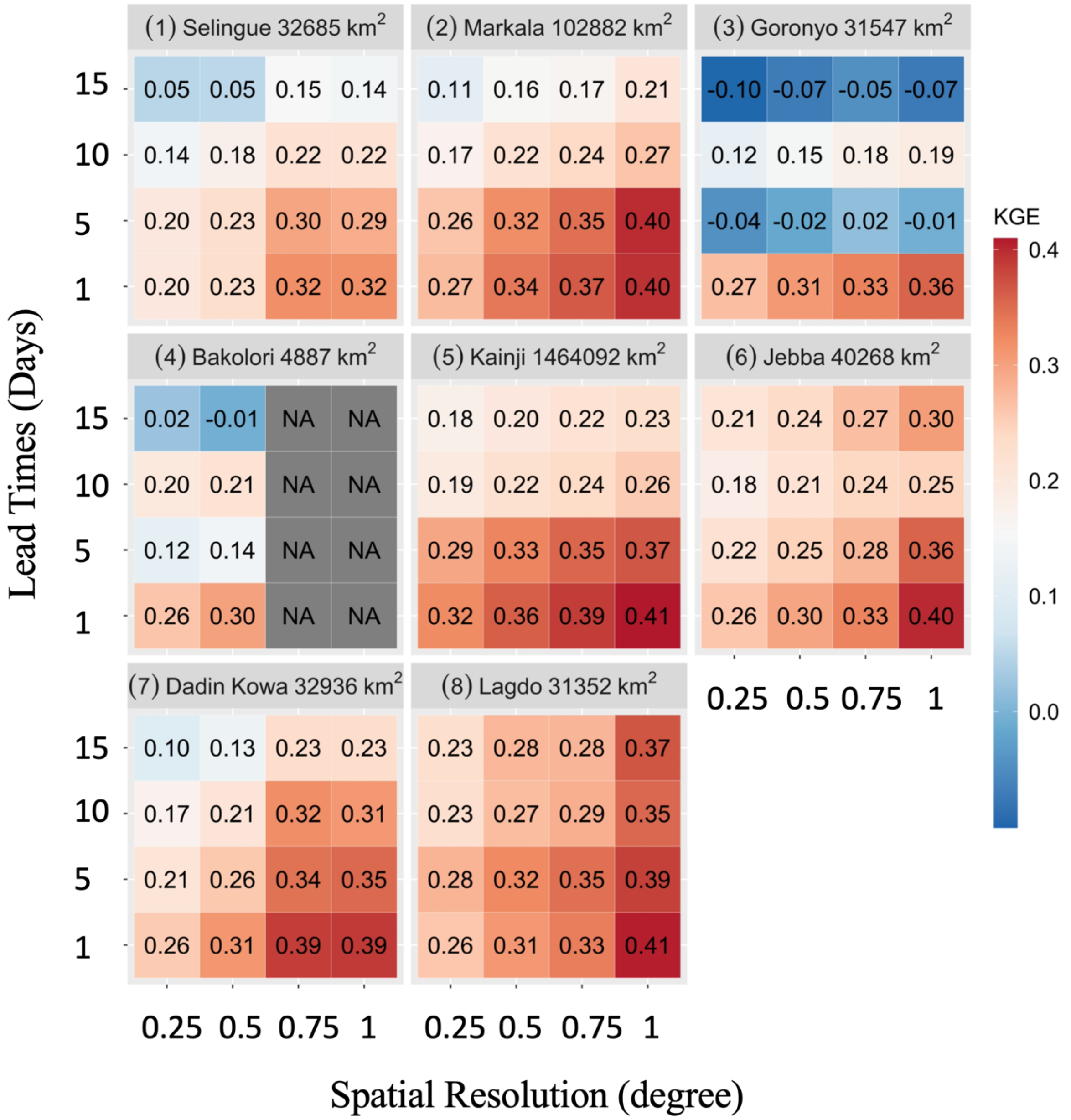


Figure 5.7. Scatterplot of watershed-averaged daily precipitation forecast obtained from 1-day lead GFS forecasts against corresponding values from IMERG Final.

5.2.5 Dependence of Daily Forecast Performance on Lead Time and Spatial Scale

In order to assess the effect of various lead times and spatial scales on forecast performance, we obtained daily GFS forecasts at various lead times (1-day, 5-day, 10-day, and 15-day), and aggregated the forecasts at spatial scales from 0.25° to coarser scales (0.5°, 0.75°, and 1°) by averaging grids. The purpose of degrading the resolution is to determine at which resolution the forecasts have acceptable performance. The KGE value at each spatial resolution was calculated in the following steps: (i) average the data at the required spatial resolution, (ii) extract pairs of data (one from IMERG Final, and the other from GFS), (iii) concatenate the pairs to form one large series of data, and (4) compute a single KGE from this data series. The resulting KGE values are shown in Figure 5.9.

With regard to the effect of spatial scales, the KGE at the GFS native resolution (i.e. 0.25°) is very low. As the spatial scale increases, KGE increases, as expected. For instance, for Markala watershed KGE increases from 0.27 (0.25°) to 0.40 (1°) for a 1-day lead. This indicates that the variation in KGE values between the watersheds could be partly explained by the watershed size. For example, based on Figure 5.5, the KGE for the 1-day lead daily GFS forecast was the highest for the largest Kainji watershed (watershed area of 1,464,092 Km2) and the lowest for the smallest Bakolori watershed (4,887 Km2). With regard to the effect of lead time for daily forecasts, KGE decreases significantly as lead time increases. For instance, for Markala watershed and a grid size of 1°, KGE decreases from 0.40 (1-day lead) to 0.21 (15-day lead).



(1) Selingue 32685 km^2

Lead time	0.25	0.5	0.75	1
15	0.05	0.05	0.15	0.14
10	0.14	0.18	0.22	0.22
5	0.20	0.23	0.30	0.29
1	0.20	0.23	0.32	0.32

(2) Markala 102882 km^2

Lead time	0.25	0.5	0.75	1
15	0.11	0.16	0.17	0.21
10	0.17	0.22	0.24	0.27
5	0.26	0.32	0.35	0.40
1	0.27	0.34	0.37	0.40

(3) Goronyo 31547 km^2

Lead time	0.25	0.5	0.75	1
15	-0.10	-0.07	-0.05	-0.07
10	0.12	0.15	0.18	0.19
5	-0.04	-0.02	0.02	-0.01
1	0.27	0.31	0.33	0.36

(4) Bakolori 4887 km^2

Lead time	0.25	0.5	0.75	1
15	0.02	-0.01	NA	NA
10	0.20	0.21	NA	NA
5	0.12	0.14	NA	NA
1	0.26	0.30	NA	NA

(5) Kainji 1464092 km^2

Lead time	0.25	0.5	0.75	1
15	0.18	0.20	0.22	0.23
10	0.19	0.22	0.24	0.26
5	0.29	0.33	0.35	0.37
1	0.32	0.36	0.39	0.41

(6) Jebba 40268 km^2

Lead time	0.25	0.5	0.75	1
15	0.21	0.24	0.27	0.30
10	0.18	0.21	0.24	0.25
5	0.22	0.25	0.28	0.36
1	0.26	0.30	0.33	0.40

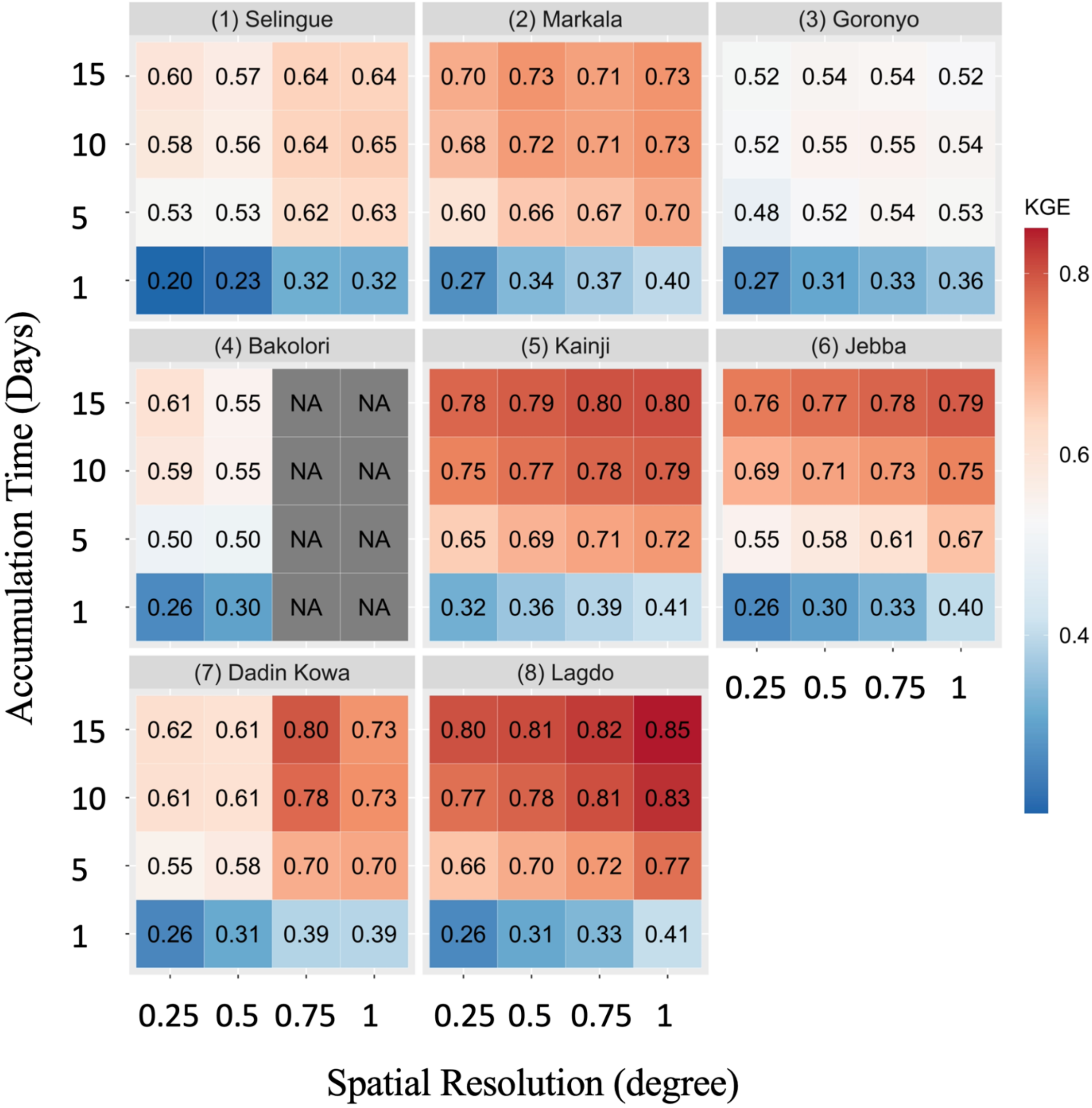
(7) Dadin Kowa 32936 km^2

Lead time	0.25	0.5	0.75	1
15	0.10	0.13	0.23	0.23
10	0.17	0.21	0.32	0.31
5	0.21	0.26	0.34	0.35
1	0.26	0.31	0.39	0.39

(8) Lagdo 31352 km^2

Lead time	0.25	0.5	0.75	1
15	0.23	0.28	0.28	0.37
10	0.23	0.27	0.29	0.35
5	0.28	0.32	0.35	0.39
1	0.26	0.31	0.33	0.41

Figure 5.8. Kling-Gupta Efficiency (KGE) for daily precipitation of GFS as a function of lead time (1-day, 5-day, 10-day, and 15-day) and spatial scale (0.25°, 0.50°, 0.75°, 1.0°). The dam names and corresponding watershed areas are given in the titles.



(1) Selingue

	0.25	0.5	0.75	1
15	0.60	0.57	0.64	0.64
10	0.58	0.56	0.64	0.65
5	0.53	0.53	0.62	0.63
1	0.20	0.23	0.32	0.32

(2) Markala

	0.25	0.5	0.75	1
15	0.70	0.73	0.71	0.73
10	0.68	0.72	0.71	0.73
5	0.60	0.66	0.67	0.70
1	0.27	0.34	0.37	0.40

(3) Goronyo

	0.25	0.5	0.75	1
15	0.52	0.54	0.54	0.52
10	0.52	0.55	0.55	0.54
5	0.48	0.52	0.54	0.53
1	0.27	0.31	0.33	0.36

(4) Bakolori

	0.25	0.5	0.75	1
15	0.61	0.55	NA	NA
10	0.59	0.55	NA	NA
5	0.50	0.50	NA	NA
1	0.26	0.30	NA	NA

(5) Kainji

	0.25	0.5	0.75	1
15	0.78	0.79	0.80	0.80
10	0.75	0.77	0.78	0.79
5	0.65	0.69	0.71	0.72
1	0.32	0.36	0.39	0.41

(6) Jebba

	0.25	0.5	0.75	1
15	0.76	0.77	0.78	0.79
10	0.69	0.71	0.73	0.75
5	0.55	0.58	0.61	0.67
1	0.26	0.30	0.33	0.40

(7) Dadin Kowa

	0.25	0.5	0.75	1
15	0.62	0.61	0.80	0.73
10	0.61	0.61	0.78	0.73
5	0.55	0.58	0.70	0.70
1	0.26	0.31	0.39	0.39

(8) Lagdo

	0.25	0.5	0.75	1
15	0.80	0.81	0.82	0.85
10	0.77	0.78	0.81	0.83
5	0.66	0.70	0.72	0.77
1	0.26	0.31	0.33	0.41

Figure 5.9. Kling-Gupta Efficiency (KGE) of GFS as a function of accumulation time scale (1-day, 5-day, 10-day, and 15-day) and spatial scale (0.25°, 0.50°, 0.75°, 1.0°).

5.2.6 Effect of Temporal Aggregation Scale on Forecast Performance

To assess the effect of temporal aggregation scale, we obtained the 1-day total, 5-day total, 10-day total, and 15-day total GFS precipitation forecasts. These multi-day forecasts are constructed by combining multiple lead-time forecasts. For instance, the 5-day total forecast is obtained by adding the 1-day lead, 2-day lead, 3-day lead, 4-day lead, and 5-day lead daily forecasts. Figure 5.10 presents the KGE values for GFS forecasts over different temporal aggregation scales, and different grid sizes. Temporal aggregation substantially increases KGE at all spatial scales. For example, at the grid size of 1° over Markala watershed, the KGE values jump from 0.40 at daily timescale to 0.73 at 15-day total timescale.

In Figure 5.11, we show the performance statistics of GFS for 15-day accumulated watershed-averaged rainfall forecast. The KGE values are intermediate ($0.5 < \text{KGE} < 0.75$) for four watersheds and good ($\text{KGE} > 0.75$) for the remaining four watersheds. Analysis of the components of KGE reveals that the improvement of KGE at longer timescales comes as a result of improved correlation and variability ratio. At the 15-day accumulation timescale, IMERG Early estimates have less bias than GFS at all watersheds, except at Kainji watershed. Figure 5.12 presents the scatterplot of 15-day accumulated GFS forecast vs IMERG Final. In general, the GFS estimates perform well for low to moderate rain rates, but tend to overestimate higher rain rates. This is consistent with Wang et al. (2019) who reported the difficulty of capturing the magnitude of high rain rates in GFS model.

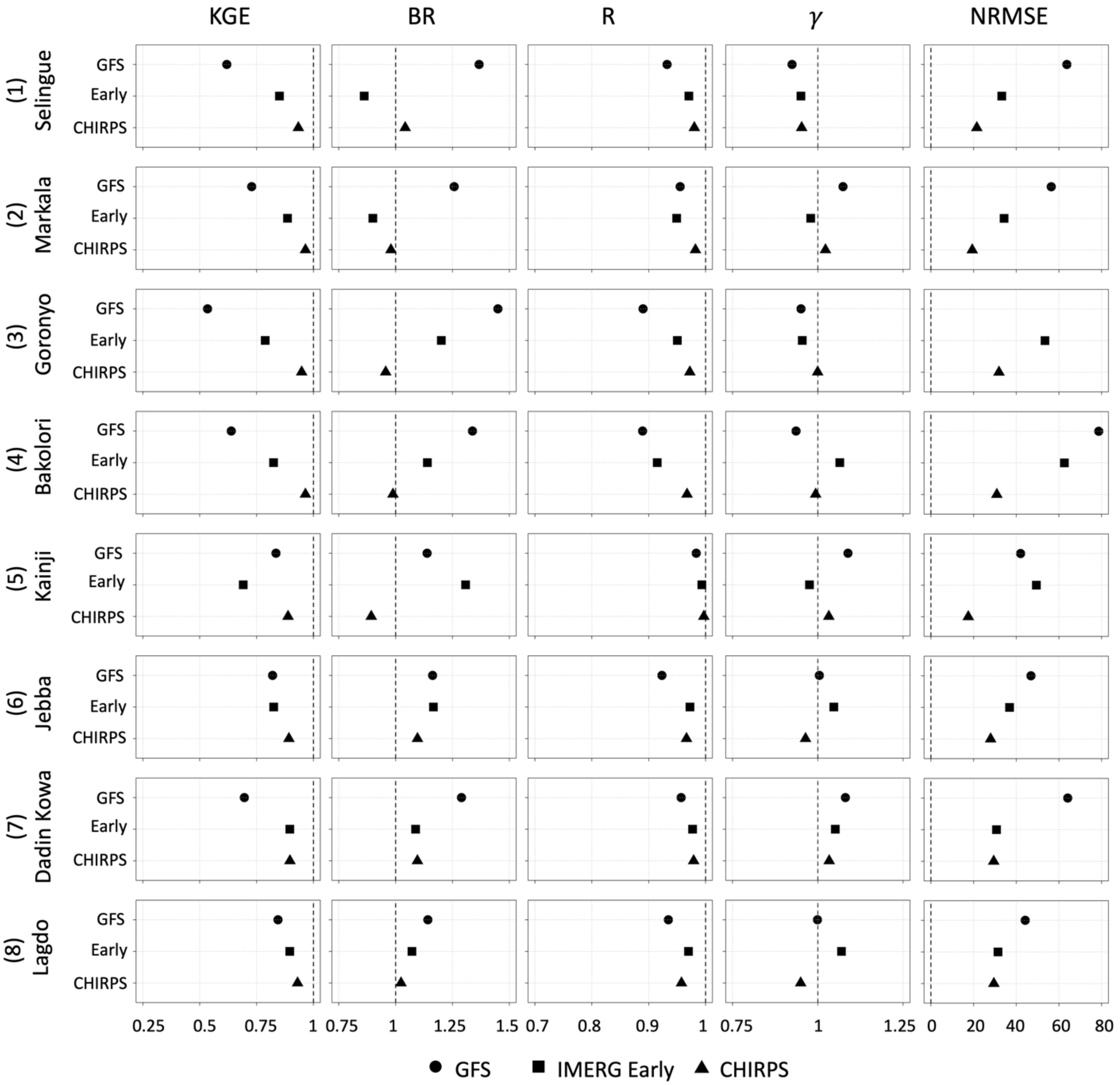


Figure 5.10. Summary of performance statistics (Kling-Gupta Efficiency KGE, Bias Ratio BR, correlation R, variability ratio γ, and root mean square error normalized by reference rainfall [%], for the 15-day accumulated GFS forecast and other satellite products.

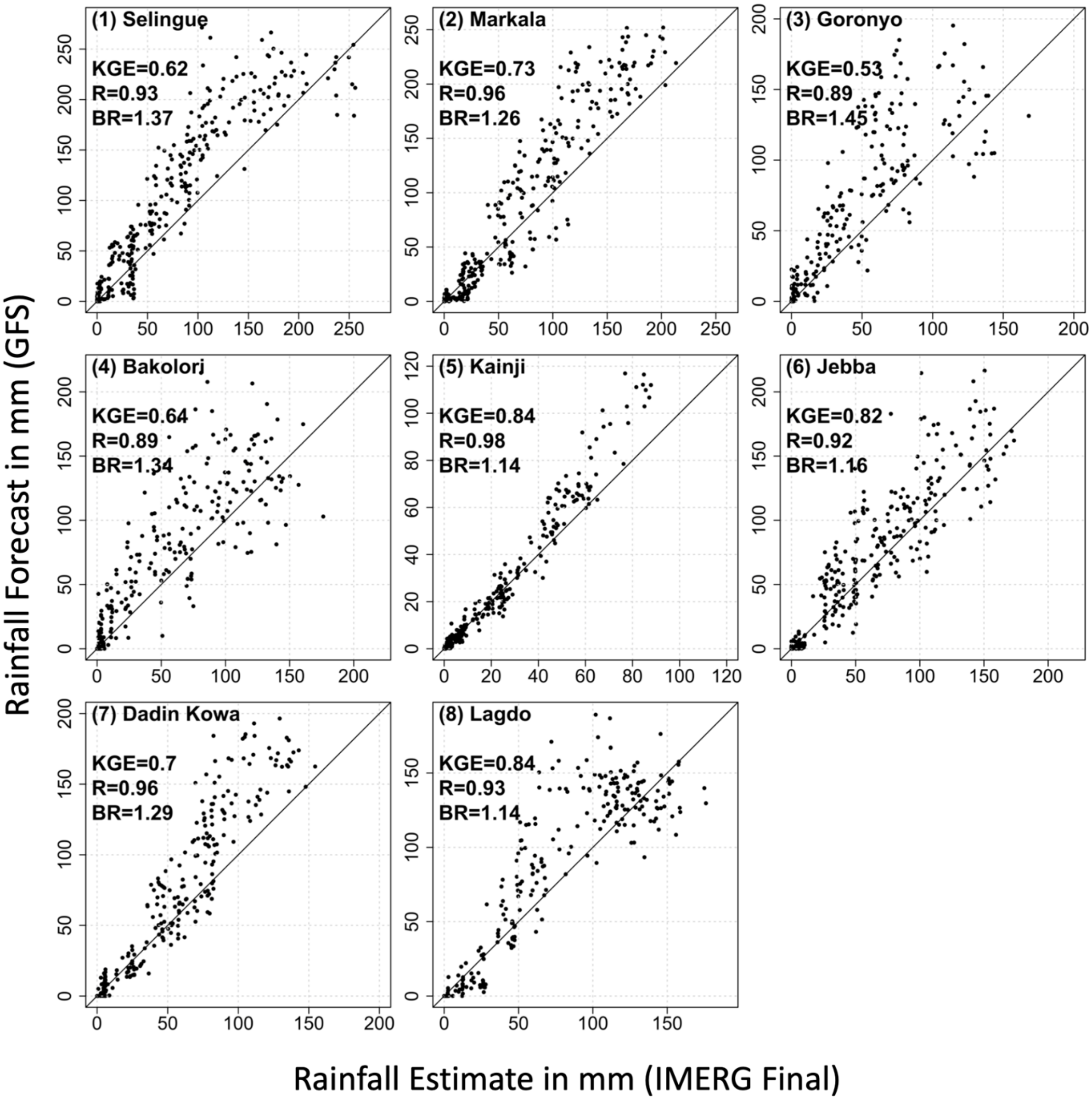


Figure 5.11. Scatterplot of watershed-averaged 15-day accumulated precipitation forecast obtained from GFS forecast against corresponding values from IMERG Final.

5.3 Conclusions

This study has evaluated the accuracy of medium-range (1-day to 15-day lead time) forecasts available from the Global Forecast System (GFS), for the watersheds of large dams in the Niger river basin. Despite the limited temporal coverage, some consistent features emerged from this evaluation. The accuracy of GFS forecast depends on climatic regime, lead time, accumulation timescale, and spatial scale. With regard to the role of climatic regimes, the GFS forecast has large overestimation bias in the Guinea (wet climatic regime), moderate overestimation bias in the Savannah (moderately wet climatic regime), but has no bias in the Sahel (dry climate). With regard to lead time, as the lead time increases, the forecast accuracy decreases. Averaging the forecasts at coarser spatial scales leads to increased forecast accuracy. For daily rainfall forecasts, the performance of GFS is very low (KGE < 0.32) at almost all watersheds except at Markala (KGE = 0.44) and Kainji (KGE = 0.68), both of which have much larger watershed areas compared to the other watersheds. Averaging the forecasts at longer time scales also leads to increased forecast accuracy. For 15-day rainfall accumulation timescale, the KGE values are either "intermediate" (i.e., 0.50 ≤ KGE ≤ 0.75) for half of the watersheds (Selingue, Goronyo, Bakolori, and Daddin Kowa) or "good" (i.e., KGE ≥ 0.75) for the remaining half (Markaa, Kainji, Jebba, and Lagdo). With regard to the effect of rainfall rate, the 15-day accumulated GFS forecasts tend to perform better for low to medium rain rates, but contain large overestimation bias at high rain rates.

The performance statistics of GFS indicate the need for calibrating GFS forecasts in order to improve their accuracy. Post-processing calibration of GFS forecasts requires the use of "relatively better performing" and "available in near-real-time" independent rainfall observations to correct real-time dynamical GFS model forecasts. This study has compared the performance of IMERG Early satellite rainfall products with the performance of GFS in terms of bias. In the Guinea and Savannah regions, IMERG Early outperforms GFS in terms of bias, while in the dry Sahel region, IMERG Early is outperformed by GFS.

We acknowledge that the reference dataset used in our evaluation (i.e., IMERG Final) has its own estimation errors. We conducted additional assessment to evaluate the performance of IMERG Final with respect to another independent and high-quality (i.e. satellite-gauge merged) rainfall product (i.e. CHIRPS). Our results show that IMERG Final and CHIRPS have similar rainfall characteristics, indicating the robustness of IMERG Final.

Overall, we conclude that the GFS forecasts, at 15-day accumulation timescale, have acceptable performance, although they tend to overestimate high rain rates. The shorter the time scale, the lower is the GFS performance. We recommend identifying suitable post-processing calibration techniques, through the use of near-real time products, such as, IMERG Early, that could improve the performance of GFS, particularly in the wet Guinea and Savannah regions. Possible calibration methods that could be explored include: simple bias (multiplicative) correction, multi-resolution bias correction through wavelet analysis wavelet analysis or empirical mode decomposition method, and Artificial-based methods such as Feed Forward Neural Network (FFNN), Support Vector Machine (SVR), and Adaptive Neural Fuzzy Inference System (ANFIS).

CHAPTER 6

Evaluation of Medium-Range Forecasts (GFS) over Senegal

6.1 Study Region

The Senegal River basin (Figure. 6.1), located in West Africa with a drainage area of 340, 000 km2, is shared by four countries: Guinea (accounting for 7% of the basin's area), Mali (35%), Mauritania (50%), and Senegal (8%) (Gaye et al. 2013). The Senegal River Valley, extending nearly 2000 km from its source in the Fouta Djalon highlands of Guinea to the Atlantic Ocean, traverses both the Sudano-Guinean and Sahelian climatic zones, and constitutes the third largest wetland ecosystem in Africa (Grosenick et al. 1990). Most of the basin has a sub-Saharan desert climate (Djaman et al. 2017). The basin has three distinct zones: the upper basin, which is mountainous, the valley, and the delta. Topographic, hydrologic, and climatic conditions are different in these regions (Djaman et al. 2017). The seasonal movement of the Inter-Tropical Convergence Zone (ITCZ) north from March to October, and south from November to February, determines seasonal and spatial variability of rainfall. The basin is home to three large dams (Table 6.1), the Manantali Dam in Mali (capacity at 11,270 Mm^3), the Foum Gleita Dam in Mauritania (500 Mm^3), and the Diama Dam in Senegal (250 Mm^3). The Manantali Dam generates hydropower (200 MW), supplies water for irrigation, and provides adequate flow for navigation on the Senegal River. The Foum Gleita Dam is primarily irrigation dam. The Diama Dam prevents saltwater intrusion upstream and supplies water for irrigation of crops. The Manantali dam, located in the mountainous and southern part of the watershed, has a drainage area of about 29,000 km2. The Foum Gleita dam, located further north with relatively low elevation, has a drainage area of about 9,500 km2. In contrast, the Diama dam, is located at the outlet of the Senegal rover basin, with ad drainage area of about 431,000 km^2.

Table 6.1. Selected dams and their watershed characteristics

Dams	Country	Operational since*	Capacity (million m3)*	Power (MW) *	Purpose*			Area of Drainage Basin (km2)**	Elevation of Drainage Basin (m)**
					Irrigation	Flood Control	Hydroelectricity		
Manantali	Mali	1988	11270	104	x		x	29340	560
Foum Gleita	Mauritania	1988	500		x			9513	128
Diama	Senegal	1986	250		x			431603	199

* information obtained from the Global Reservoir and Dam Database (Lehner et al. 2011) and Food and Agriculture Organization of the United Nations (FAO)'s Global Information System on Water and Agriculture (AQUASTAT).
** Calculated from HydroSEHDS (Lehner et al. 2008).

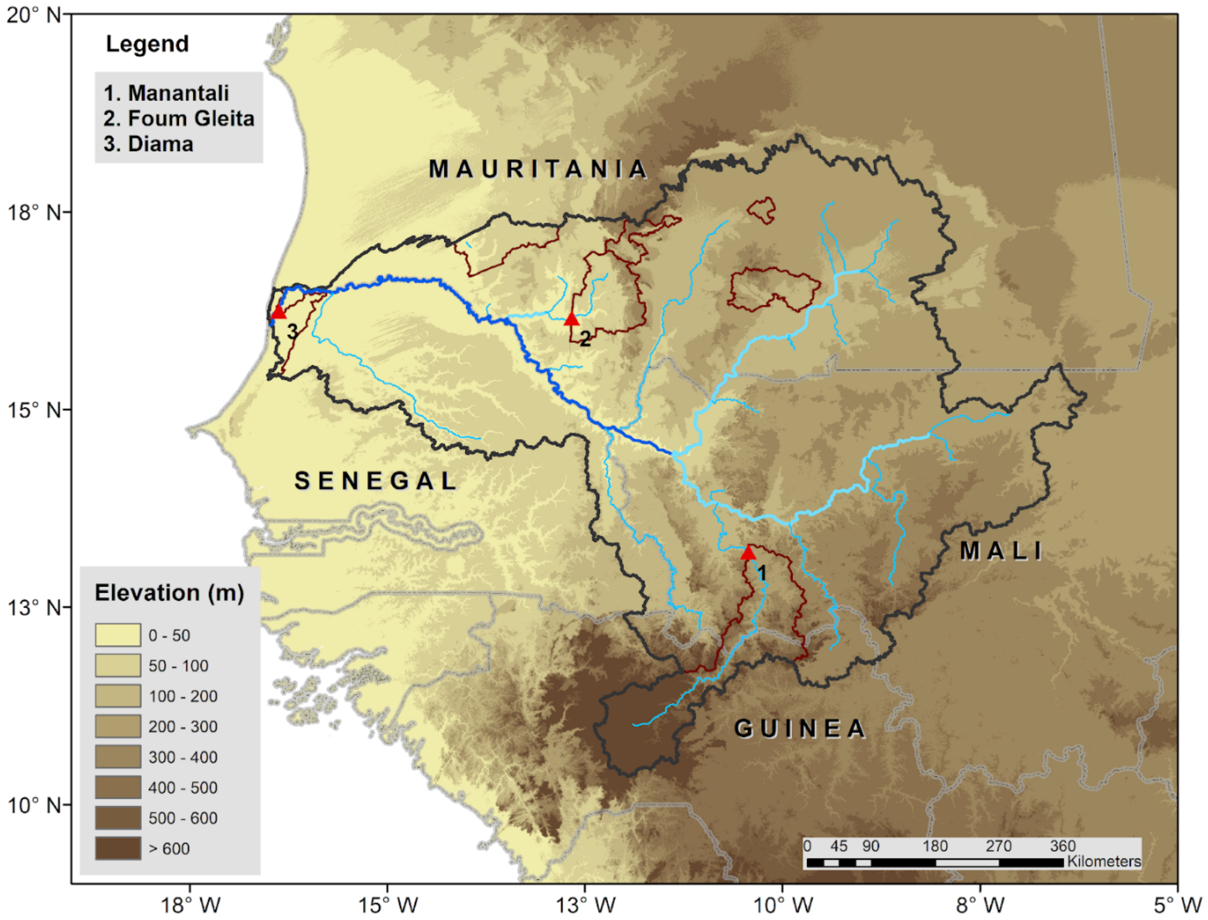


Figure 6.1 Map of Senegal River Basin, with the location of selected dams/reservoirs: (1) Manantali, (2) Foum Gleita, and (3) Diama, and the drainage basins defined by the dam locations.

6.2 Results and Discussion

6.2.1 Annual Spatial Variability of Rainfall

Figure 6.2 presents the spatial map of annual (15 June 2019 to 15 June 2020) rainfall derived from the 1-day lead GFS forecast and satellite products. According to the reference IMERG Final, the annual rainfall shows a distinct south-north gradient, varying from more than 1,500 mm in the southern part (Savannah climate, and mountainous) to under 250 mm in the northern part (Sahel climate, and low-elevation). Validated against IMERG Final, the 1-day lead GFS captures well the spatial pattern of rainfall. The correlation between the spatial distribution of annual rainfall derived from the 1-day lead GFS and IMERG Final is 0.94.

6.2.2 Monthly Cycle of Rainfall

Figure 6.3 presents the time series of monthly rainfall derived from the 1-day lead GFS and satellite precipitation products. According to IMERG Final, rainfall in the region has one rainy season with a peak in August in all three watershed cases. The main mechanism for this precipitation is the northward migration of the Inter Tropical Convergence Zone (ITCZ) during summer. The rainy season in the Foum Gleita watershed (dry Sahel climate) is of short period (July through October), with an amplitude of 100 mm month^{-1}, whereas the rainy season over the Manantali watershed (wet Savannah climate) is of longer period (May through October) with an amplitude of around 400 mm month^{-1}. The 1-day lead GFS captures well the seasonal cycle of rainfall in all three watershed cases but tends to overestimate the monthly rainfall over Manantali, slightly underestimate over Foum Gleita, and is almost unbiased over the larger Diama watershed, which integrates both the wet and dry parts of the watershed.

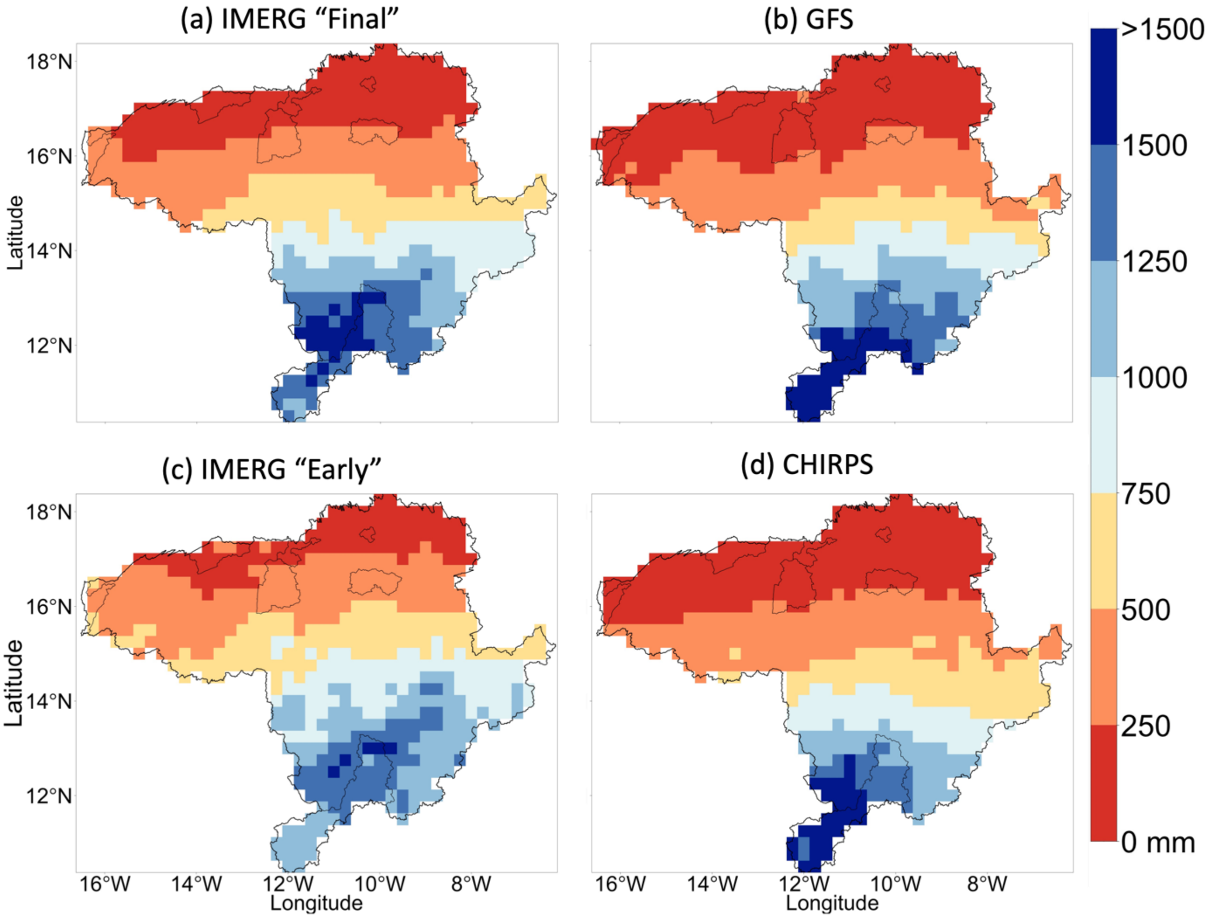


Figure 6.2. Spatial map of annual rainfall (in mm), for the period 15 June 2019 to 15 June 2020, derived from (a) IMERG Final, (b) GFS (1-day lead time), (c) IMERG Early, and (d) CHIRPS.

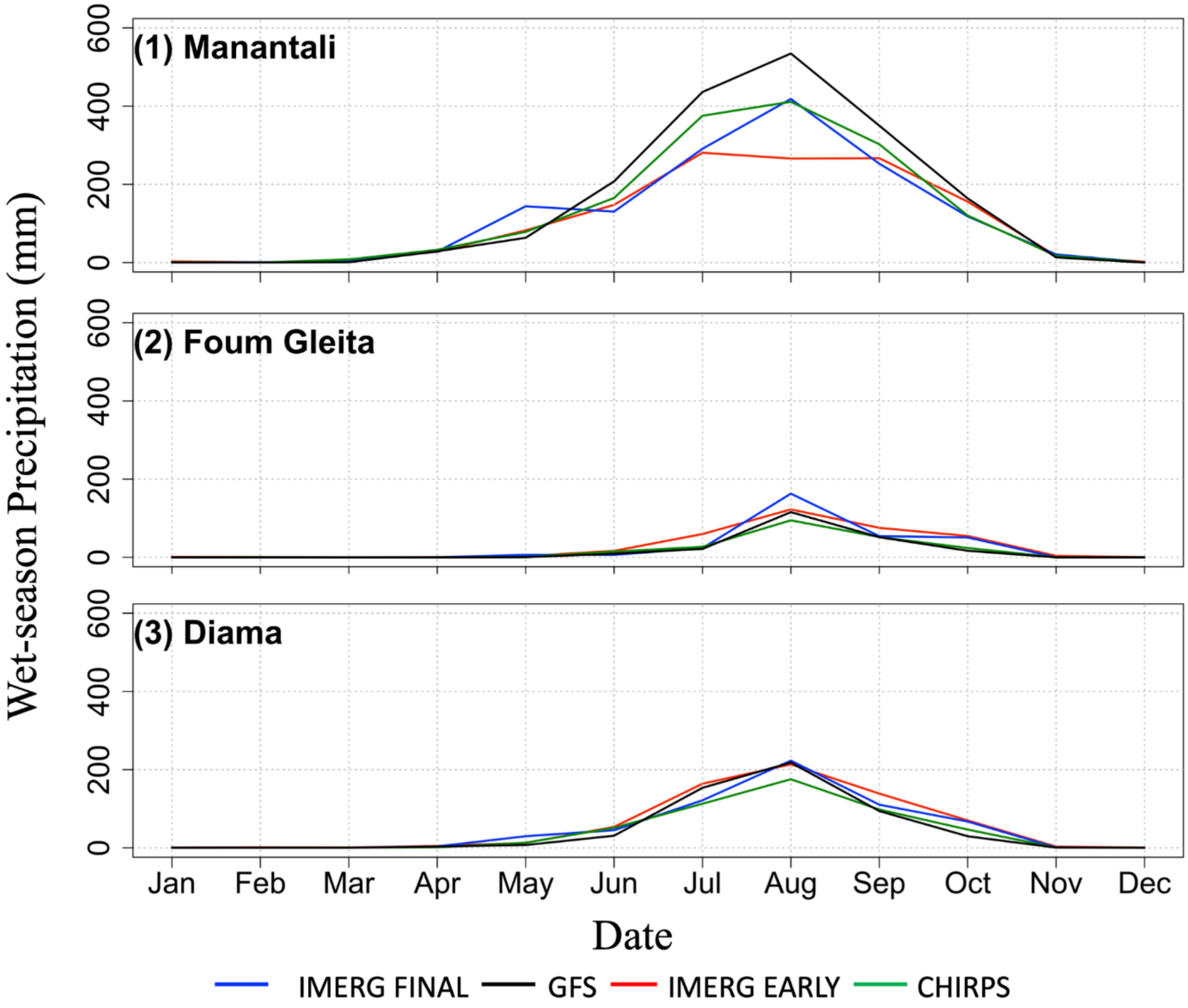


Figure 6.3 Monthly time series of sub-basin averaged precipitation (mm), for the period 15 June 2019 to 15 June 2020, derived from IMERG Final, GFS (1-day lead time), IMERG Early, and CHIRPS, for watersheds of three dams in the Senegal River Basin: (1) Manantali, (2) Foum Gleita, and (3) Diama.

6.2.3 Annual Rainfall

Here, we aggregate the 1-day lead GFS forecasts to annual time scale and compare the results against corresponding annual precipitation estimates from IMERG Final (Figure 6.4). According to IMERG Final, the annual watershed-average rainfall over the dam watersheds is: 300 mm (Foum Gleita), 600 mm (Diama), and 1400 mm (Manantali). The 1-day lead GFS overestimates the rainfall over Manantali by 28%, while it slightly underestimates rainfall over Foum Gleita (by 30%) and Diama (by 11%).

6.2.4 Daily Time series

Figure 6.5 presents the daily time series of watershed-averaged rainfall, derived from 1-day lead GFS and satellite products. According to IMERG Final, the daily time series of rainfall is characterized by several spikes in all three watersheds. Over the Manantali and Diama watersheds, there is rainfall occurrence on a daily basis, whereas over Foum Gleita, there are no-rain periods during the rainy season. The GFS captures well the daily timeseries rainfall pattern over all three dam watersheds; however, GFS tends to overestimate the majority of daily rainfall over Manantali, underestimate spikes over Foum Gleita, and estimates well the magnitudes of daily rainfall over Diama.

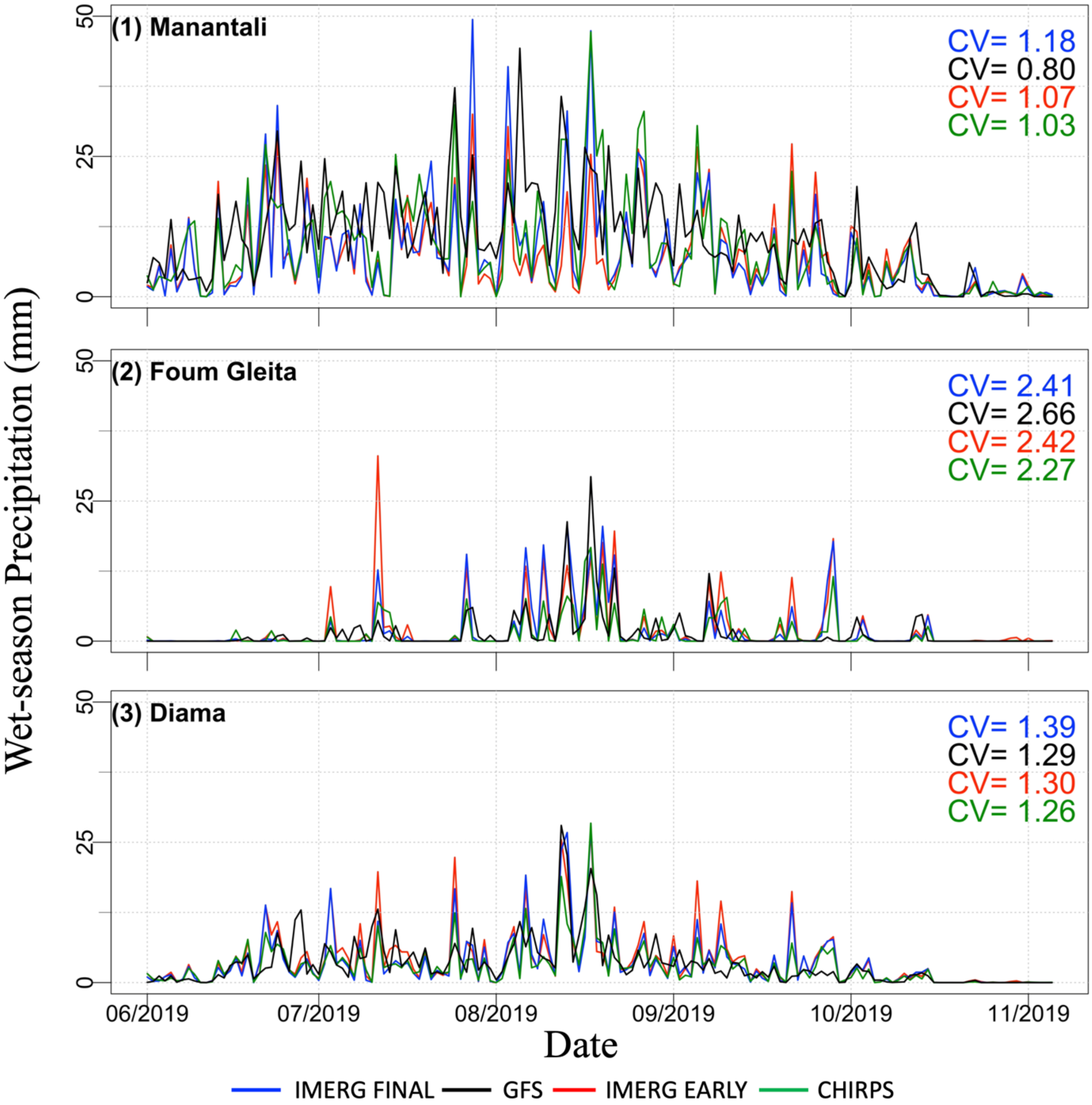


Figure 6.5. Time series of sub-basin averaged precipitation total (mm) for the wet period (November – April), for each of the dam watersheds, as derived from various precipitation products. The Figure also shows the coefficient of variation (CV) as a measure of temporal variation.

6.2.5 KGE Statistics

Figure 6.6 presents the performance statistics of various precipitation products (with respect to the IMERG Final estimates), in terms of Kling-Gupta Efficiency (KGE), Bias Ratio (BR), correlation (R), variability ratio (γ), and root mean square error normalized by reference precipitation mean (NRMSE). The KGE of the 1-day lead, daily, GFS forecast is 0.29 (Manantali), 0.53 (Foum Gleita), and 0.65 (Diama). The GFS skill is therefore "poor" over the mountainous and wet watershed (Manantali), while it is "intermediate" over the dry watershed (Foum Gleita) and the watershed with the largest area (Diama). The low KGE for Manantali can be attributed to low performances in all components of KGE: high bias (BR = 1.37), low correlation (R = 0.50), and low variability ratio (γ = 0.68), indicating that the GFS over the wet and mountainous part of the watershed is characterized by large systematic and random error. For the Foum Gleita watershed, the KGE value (0.53) is mainly affected by both bias (BR = 0.75) and low correlation (R = 0.62). For the Diama watershed, the KGE value (0.65) is affected mainly by the low correlation (R = 0.66). The root-mean-square-error of the GFS varies in the range 100% to 300% of the mean rainfall.

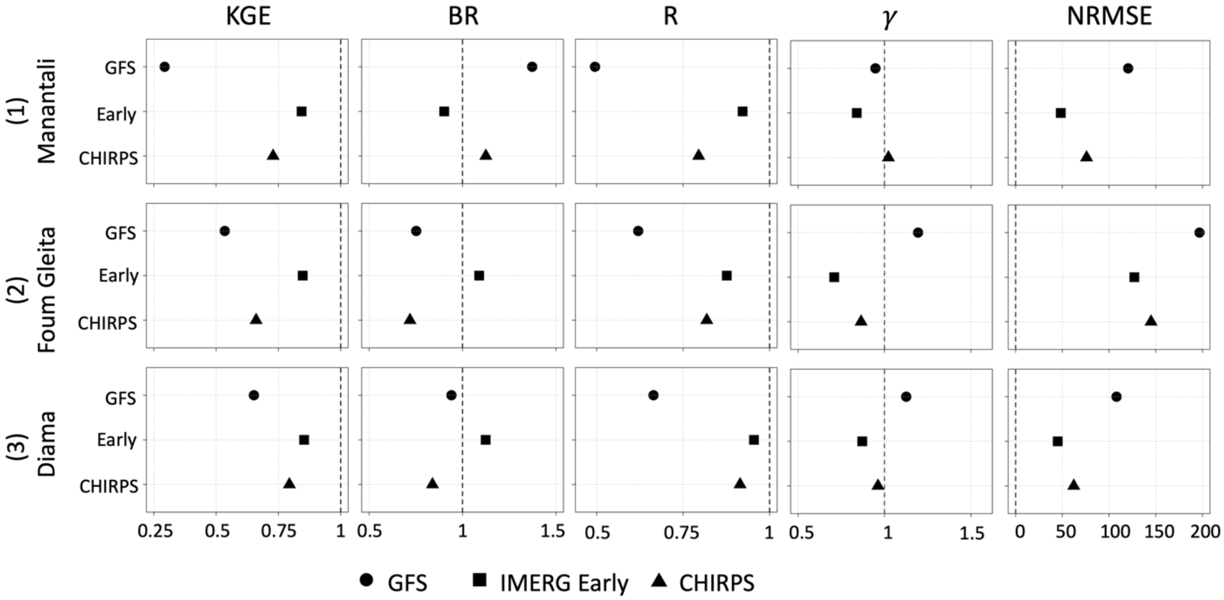


Figure 6.6. Summary of performance statistics (Kling-Gupta Efficiency KGE, Bias Ratio BR, correlation R, variability ratio γ, and root mean square error normalized by reference precipitation mean NRMSE) of 1-day lead GFS forecasts and different satellite products, during the wet periods, at different watersheds.

6.2.6 Dependence of Forecast Performance on Precipitation Rate

Figure 6.7 presents the scatterplot of 1-day lead GFS forecasts against daily IMERG Final rain rates. At the wet Manantali watershed, the GFS has false alarms, and tends to overestimate light rainfall and underestimate heavy rainfall. At the dry Foum Gleita watershed, the GFS captures well light rainfall, but tends to heavily underestimate heavy rainfall. At the large Diama watershed, there is between agreement between GFS and IMERG Final, but still with some overestimation of light rainfall and some underestimation of heavy rainfall.

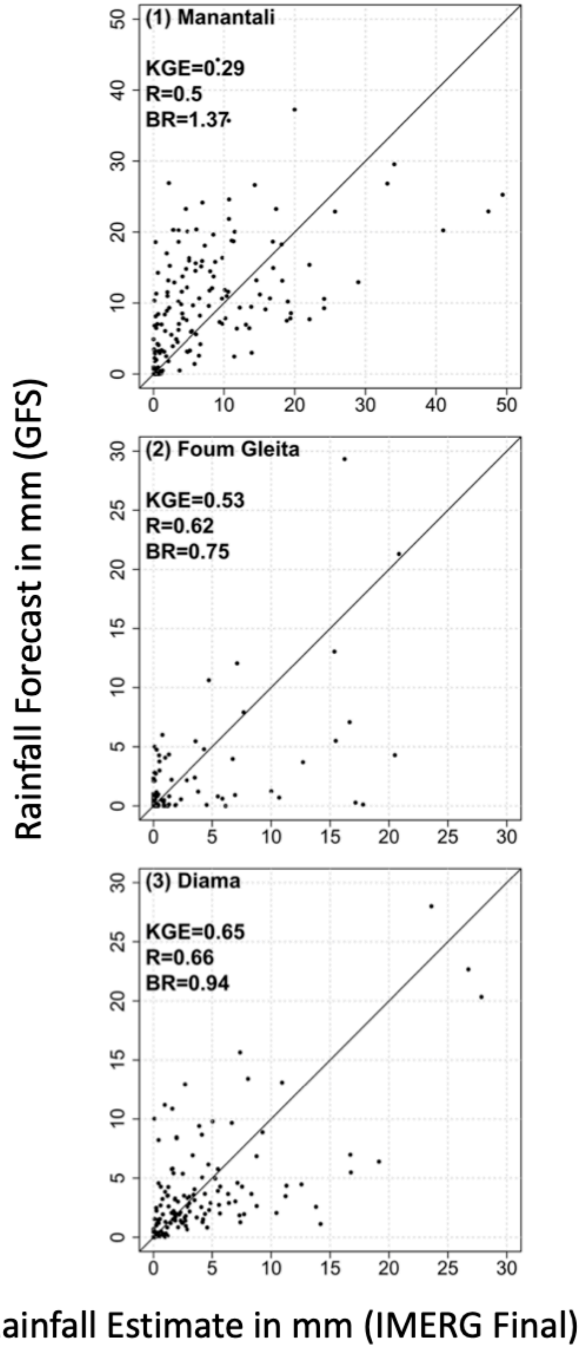


Figure 6.7 Scatterplot of watershed-averaged daily precipitation forecast obtained from 1-day lead GFS forecasts against corresponding values from IMERG Final, over each dam watershed in the Senegal River Basin.

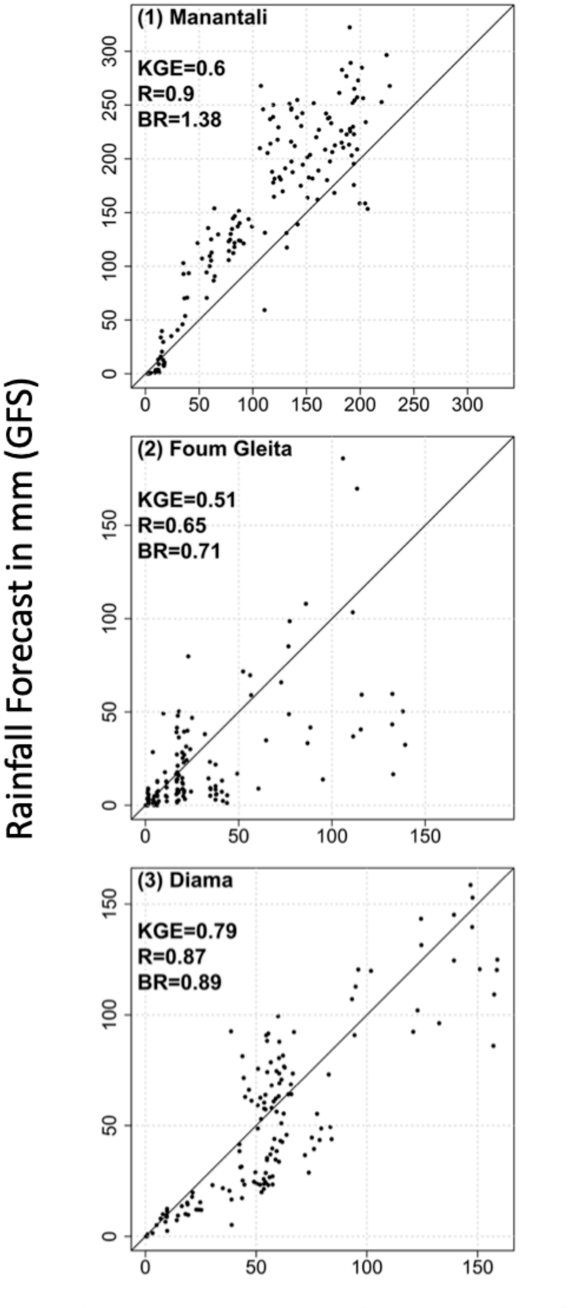


Figure 6.8 Same as for Figure 7 but for 15-day accumulation.

6.2.7 Effect of Temporal Aggregation Scale on Forecast Performance

Figure 6.8 presents the Kling-Gupta Efficiency (KGE) and its components for 15-day accumulated GFS rainfall forecast. We obtained the 15-day accumulated forecast by summing up the daily forecasts at various lead times, starting from 1-day lead all the way to 15-day lead. Aggregating the forecasts from daily forecasts to 15-day accumulated forecasts has improved the performance of GFS for the wet Manantali watershed and the larger Diama watershed, whereas there is no much improvement for the dry watershed, Foum Gleita. Over Manantali watershed, the 15-day accumulated forecast has shown increases in KGE (0.60 compared to 0.29 at daily timescale) and correlation (0.90 compared to 0.50 at daily), but no change in bias ratio (BR = 1.38). Over the Diama watershed, the 15-day accumulated rainfall has similarly shown increases in KGE (0.79 compared to 0.65 at daily) and correlation (0.87 compared to 0.66 at daily).

6.2.8 Comparison of the Performances of IMERG Early and GFS

As discussed above, the skills of 1-day lead, daily, GFS forecasts are poor for Manantali but medium for Foum Gleita and Diama Dams. One of the possible methods to improve GFS forecast accuracy is post-processing of GFS using rainfall estimates that have relatively better accuracy and are available in near-real time. Satellite-only products, such as IMERG Early, are available in near-real time. In this section, we explore how the performance of IMERG Early compare with the performance of GFS. Like GFS, IMERG Early captures the spatial pattern of annual rainfall (Figure. 6.2). Like GFS, IMERG Early captures the seasonal cycle of rainfall (Figure. 6.3). Over the wet and mountainous Manantali watershed, IMERG Early captures well the monthly rainfall except for August, where IMERG Early underestimates. Over the other two watersheds (Foum Gleita and Diama), IMERG Early has similar performance with GFS in terms of capturing monthly time series. At the annual timescale, over the Manantali watershed (where

GFS overestimates by 37%), IMERG Early underestimates by 10% (Fig. 4). Over the dry Foum Gleita watershed, IMERG Early is almost identical to IMERG Final, whereas GFS underestimates annual rainfall by 30%. Over the large Diama watershed, IMERG Early overestimates by 10%, whereas GFS underestimates by 11%. The KGE of IMERG Early is above 0.75 (i.e. “good skill”) and is higher than that of GFS at all three watersheds (Fig. 6). IMERG Early outperforms GFS in all components of KGE, that is, bias, correlation, and variability ratio. The root-mean-square-error of IMERG Early is only about 50% of the mean at the Manantali and Diama watersheds but is high at about 127% at the dry Foum Gleita watershed. Therefore, our analysis indicates that IMERG Early has better capabilities than GFS, and is therefore worthy of consideration as input into post-processing techniques aimed at improving the accuracy of GFS forecasts,

6.2.9 Performance of GFS if the Reference Product is Changed from IMERG Final to CHIRPS

We acknowledge that the reference dataset used in our evaluation (i.e., IMERG Final) has its own estimation errors. We conducted additional assessment to evaluate the performance of GFS using CHIRPS rainfall products as reference. Table 6.2 shows the performance statistics of GFS for different lead times, using IMERG FINAL and CHIRPS, separately, rainfall products as reference. The overall magnitude of GFS performance (as well as the variability of the performance across watersheds) is similar when either rainfall product is used as a reference. Both reference products reveal that the GFS forecast has relatively low skill (and very high bias, and very low correlation) in the wet and mountainous region of Manantali dam watersheds and the skill improves over the dray watershed of Foum Gleita and the larger watershed of Diama. Therefore, our results show that the overall performance of GFS remains the same if the reference product were to be changed from IMERG Final to CHIRPS, indicating the robustness of IMERG Final as reference product.

Table 6.2 Performance statistics of daily GFS forecast for various lead times (1-day, 5-day, 10-day, and 15-day) using IMERG Final (CHIRPS) rainfall products as reference, in terms of correlation, bias ratio, and NRMSE.

Lead time of GFS forecast	**Correlation**	**Bias Ratio**	**KGE**	**NRMSE (%)**
		Manantali		
1-day	0.50 (0.56)	1.37 (1.22)	0.29 (0.46)	120.43 (96.36)
5-day	0.43 (0.44)	1.22 (1.09)	0.31 (0.39)	119.28 (102.45)
10-day	0.23 (0.41)	1.39 (1.24)	0.07 (0.31)	144.50 (110.58)
15-day	0.09 (0.20)	1.61 (1.44)	-0.13 (0.07)	183.55 (150.01)
		Foum Gleita		
1-day	0.62 (0.58)	0.75 (1.05)	0.53 (0.55)	196.81 (234.39)
5-day	0.34 (0.44)	0.67 (0.93)	0.06 (0.12)	287.39 (326.90)
10-day	0.22 (0.29)	0.78 (1.08)	0.19 (0.28)	273.15 (295.28)
15-day	0.42 (0.30)	0.70 (0.97)	0.33 (0.27)	237.56 (293.25)
		Diama		
1-day	0.66 (0.68)	0.94 (1.12)	0.65 (0.66)	108.01 (109.38)
5-day	0.60 (0.67)	0.81 (0.96)	0.55 (0.65)	116.49 (104.98)
10-day	0.36 (0.44)	0.94 (1.12)	0.34 (0.42)	143.57 (136.92)
15-day	0.31 (0.30)	0.93 (1.10)	0.30 (0.29)	148.46 (152.85)

6.3 Conclusions

The objective of this study was to evaluate the accuracy of medium-range (1-day to 15-day lead time) forecasts available from the Global Forecast System (GFS) in the Senegal River Basin, focusing on some of its major dams, namely, Manantali, Foum Gleita, and Diama. The watershed of Manantali dam is mountainous and located in a wet climate, with annual rainfall of 1400 mm during the study period. Foum Gleita, on the other hand is located in the dry Sahel, with annual rainfall of 300 mm during the study period and has a low-elevation terrain. Diama is a very large watershed and covers the entire basin, with annual rainfall of 600 mm during the study period. The evaluation is done using the satellite-gauge merged rainfall product IMERG

Final as reference, and additional analysis is conducted by changing the reference product to CHIRPS. Additionally, the performance of GFS is compared to the performance of the near-real time, satellite-only rainfall product, IMERG Early, to see if IMERG Early has the potential to be used in post-processing of GFS.

The results show that the 1-day lead GFS has lower performance in the mountainous and wet region of Manantali (KGE = 0.29) and relatively higher performance in the low-elevation and dry areas of Foum Gleita and Diama (KGE = 0.53 to 0.59). The lower KGE over Manantali is due to the high overestimation bias (overestimation by 37%) in the GFS forecasts and low correlation coefficient (R = 0.50) between the daily time series of 1-day lead GFS forecast and IMERG Final.

Aggregating the forecasts from daily to 15-day accumulation increases the performance of the GFS forecasts. The KGE of the 15-day accumulated forecast showed “intermediate” skill (KGE = 0.60) for Manantali, and “good” skill (KGE = 0.79 and 0.87) for Foum Gleita and Diama watersheds. IMERG Early outperforms GFS in all the watershed cases. In particular, IMERG Early results in lower bias (underestimation by 10%) compared to the large overestimation bias (overestimation by 37%) over the Manantali watershed.

Therefore, we conclude that the GFS forecasts in the Senegal River Basin have good skills at large temporal aggregation scales (~ 15-day), but their short-lead forecasts need further improvement before they can be used in applications. Given that IMERG Early outperforms GFS, we recommend testing the suitability of IMERG Early to serve as input into post-processing of GFS in order to improve the accuracy of GFS forecasts. Possible post-processing techniques that could be explored include: simple bias (multiplicative) correction (Gumindoga et al. 2019), multi-resolution bias correction through wavelet analysis (Xu et al. 2019) or empirical mode decomposition method (Wang et al. 2020, Prasad et al. 2019), and Artificial Intelligence-based methods such as Feed Forward Neural Network (Cloud et al. 2019), Support Vector Machine ((Du et al. 2017; Yu et al. 2017), and Adaptive Neural Fuzzy Inference System (Jehanzaib et al. 2021).

CHAPTER 7

Evaluation of Medium-Range Forecasts (GFS) over Volta

7.1 Study Region

The Volta River basin (Figure 7.1) in West Africa covers an area of 409,000 km2 and is shared by six riparian countries where Ghana and Burkina Faso make up the largest area of 42% and 43%, respectively, while the remaining 15% is distributed between Benin, Côte d'Ivoire, Mali and Togo (Mul et al. 2015; Williams et al. 2016). The physical relief of the basin is generally low with predominantly undulating topography except for the Akwapim-Togo mountain ranges located in the eastern part. The climate of the basin is strongly influenced by the movement of the Inter-Tropical Convergence Zone (ITCZ) which produces low rainfall in the northern part of the basin and large rainfall in the southern part of the basin (Baah-Kumi & Ward. 2020).

The Volta River system consists of the Black Volta in the west, the White and Red Volta (each flowing from Burkina Faso), and the Oti River in the east (draining the highlands of Togo and Benin), all of which join in Ghana to form the lower Volta before flowing into the Gulf of Guinea. The Akosombo dam on the Lower Volta formed the Lake Volta, one of the largest man-made reservoirs in the world, covering 8,500 km^2, or 4% of Ghana's land area (Barry et al. 2021), with a gross storage capacity of 148 Bm3 and an active capacity of 78 Bm^3. Akosombo Dam, with a drainage area of 407,093 km^2, is a hydropower dam that generates 1012 MW, of which around 20% is supplied to the national grid and the remaining 80% is supplied to the aluminum industry.

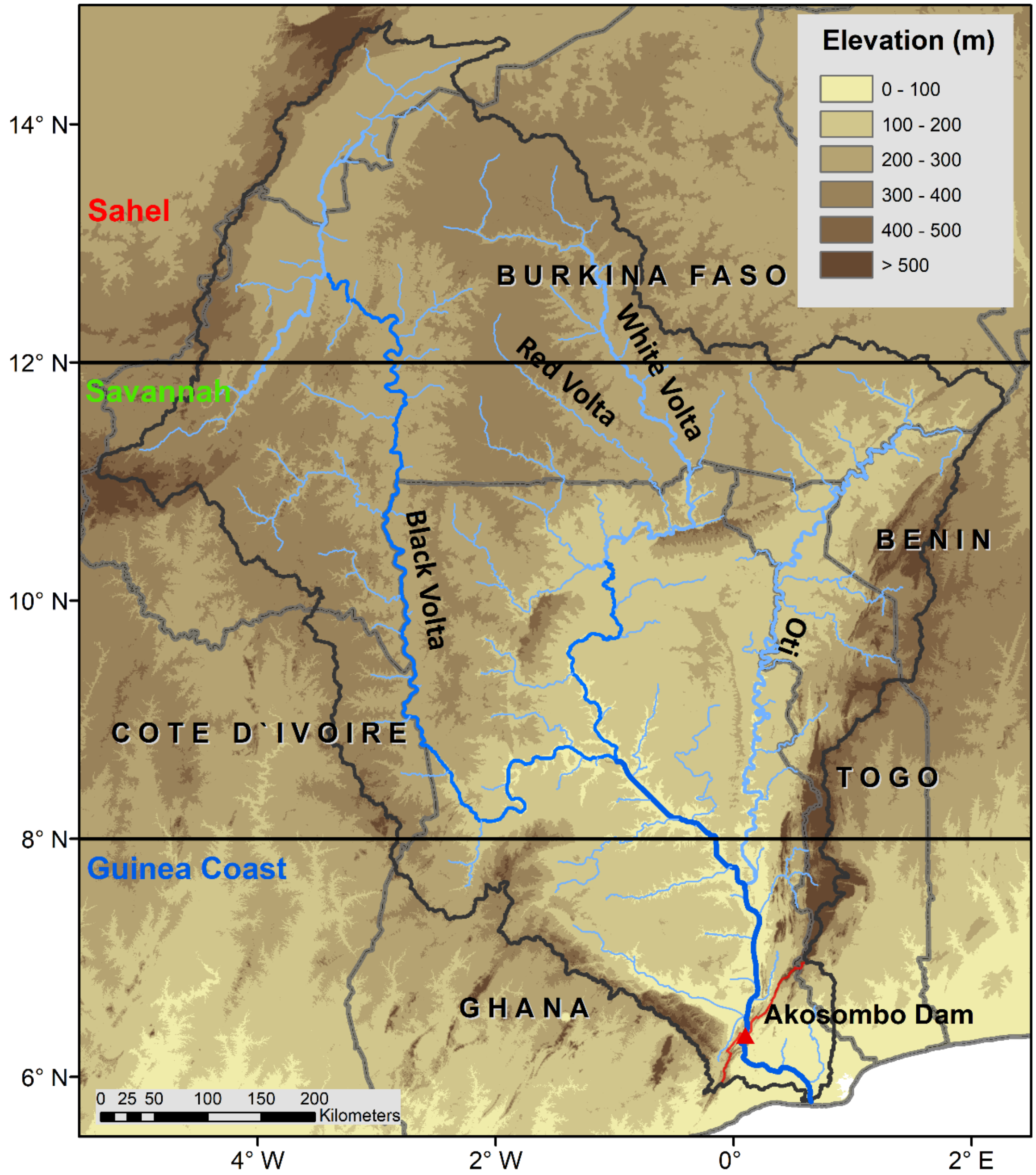


Figure 7.1. Map of the Volta River Basin, with the location of the Akosombo dam.

7.2 Results and Discussion

7.2.1 Annual Spatial Variability and Seasonal Characteristics

Figure 7.2 presents the spatial map of annual (15 June 2019 to 15 June 2020) rainfall derived from GFS and three satellite precipitation products. According to the reference rainfall product (i.e. IMERG Final), the spatial rainfall pattern is characterized by a strong north-south gradient: 815 mm yr^{-1} in the Sahel to 1,190 mm yr^{-1} in the Savannah to 1,300 mm yr^{-1} in the Guinea Coast zone. The 1-day lead GFS forecast captures the overall spa-tial pattern, but with large underestimation in the dry Sahel (656 mm yr^{-1} on average) and large overestimation in the Guinea Coast (1,300 mm yr^{-1} on average), while producing a relatively more accurate result in the Savannah (1,130 mm yr^{-1} on average).

The uncalibrated IMERG Early estimates capture the overall spatial pattern, but result in large overestimation in the Sahel (1,000 mm yr^{-1} on average), while providing more accu-rate results in both Savannah and Guinea Coast. Comparison of the performance of GFS and IMERG Early, when both are validated against IMERG Final, show that the GFS out-performs IMERG Early in the Guinea Coast, but both provide highly biased estimates in the dry Sahel region with contradicting bias characteristics (i.e. large negative bias by GFS, but large positive bias by IMERG Early). The CHIRPS estimates capture well not only the overall spatial pattern, but also the actual magnitudes.

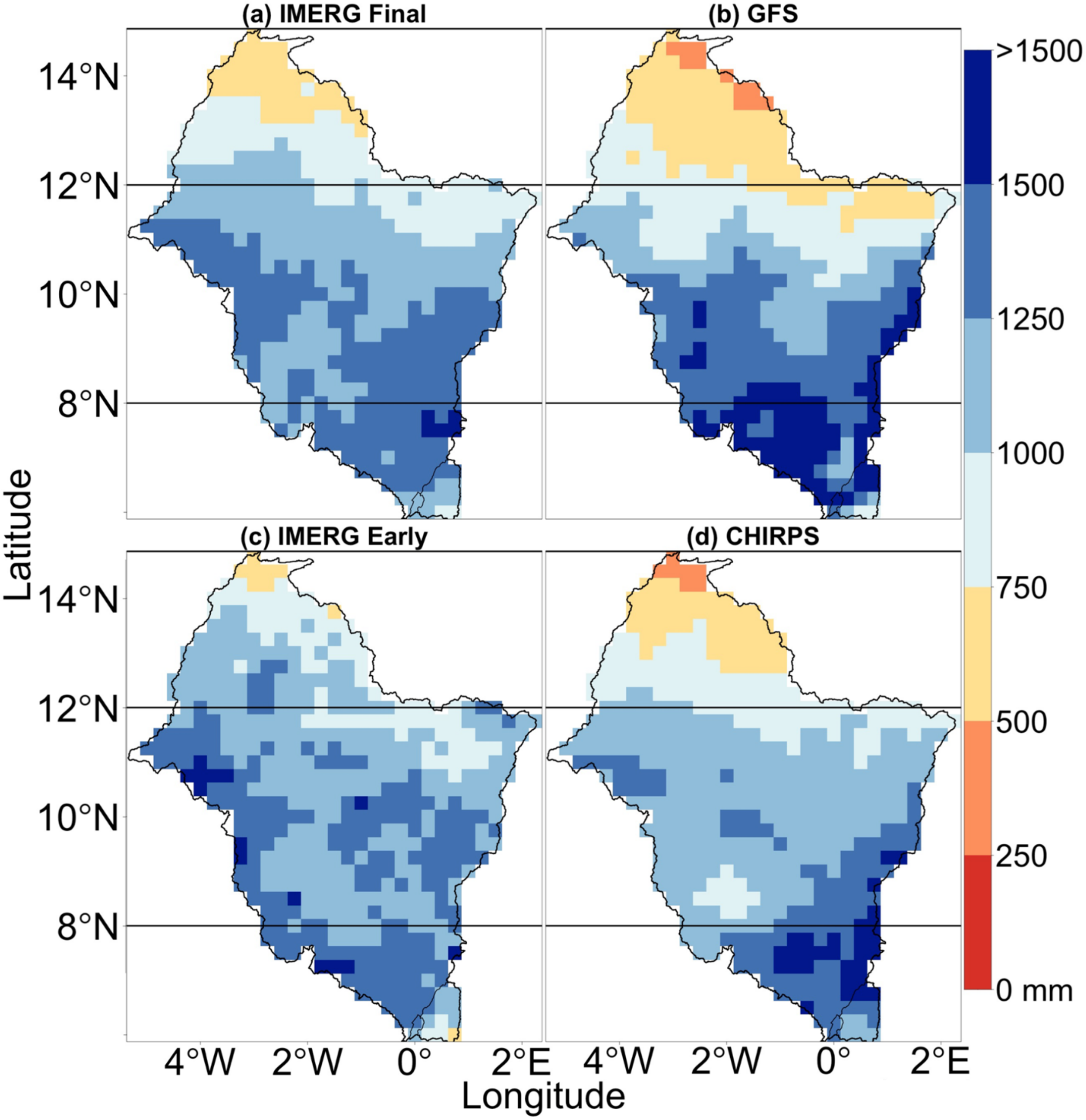


Figure 7.2. Spatial map of annual rainfall (in mm), for the period 15 June 2019 to 15 June 2020, derived from (a) IMERG Final, (b) GFS (1-day lead time), (c) IMERG Early and (d) CHIRPS.

Figure 7.3 presents the monthly time series of climatological-zone averaged rainfall derived from the 1-day lead GFS and three satellite precipitation products. Let us look at results in the Sahel region. According to the reference rainfall (i.e. IMERG Final), the rainy season is six months long, from May until November. The rainfall peak is about 220 mm/month during July and August. The 1-day lead GFS forecast shows similar seasonal pattern but tends to underestimate in all months except in August. In contrast, IMERG Early is close to the reference product in almost all months but has large overestimation bias in two months (August and September). The CHIRPS estimates capture well the seasonal pattern as well as the actual magnitude of rainfall. In the Savannah region, the rainfall starts early in March, and large rainfall accumulations occur over four months (July to October). In the first half of the rainy season (March through July), GFS underestimates rainfall, whereas in the second half, GFS captures well the monthly rainfall values, and GFS overestimates in August. In contrast, IMERG Early provides more accurate results in the first half of the rainy season including August, but it overestimates rainfall in September and October. Therefore, in the first half of the rainy season including August, IMERG Early outperforms GFS, but in the second half GFS outperforms IMERG Early. The CHIRPS estimates are al-most identical to the IMERG Final estimates. Guinea Coastal zone is characterized by similar rainfall season as in Savannah but with large rainfall accumulations starting from as early as March. GFS performs well in the first half of the rainy season but tends to over-estimate in the second half. In contrast, IMERG Early performs well in almost all months, except for October when it overestimates. CHIRPS monthly estimates are almost identical to IMERG Final.

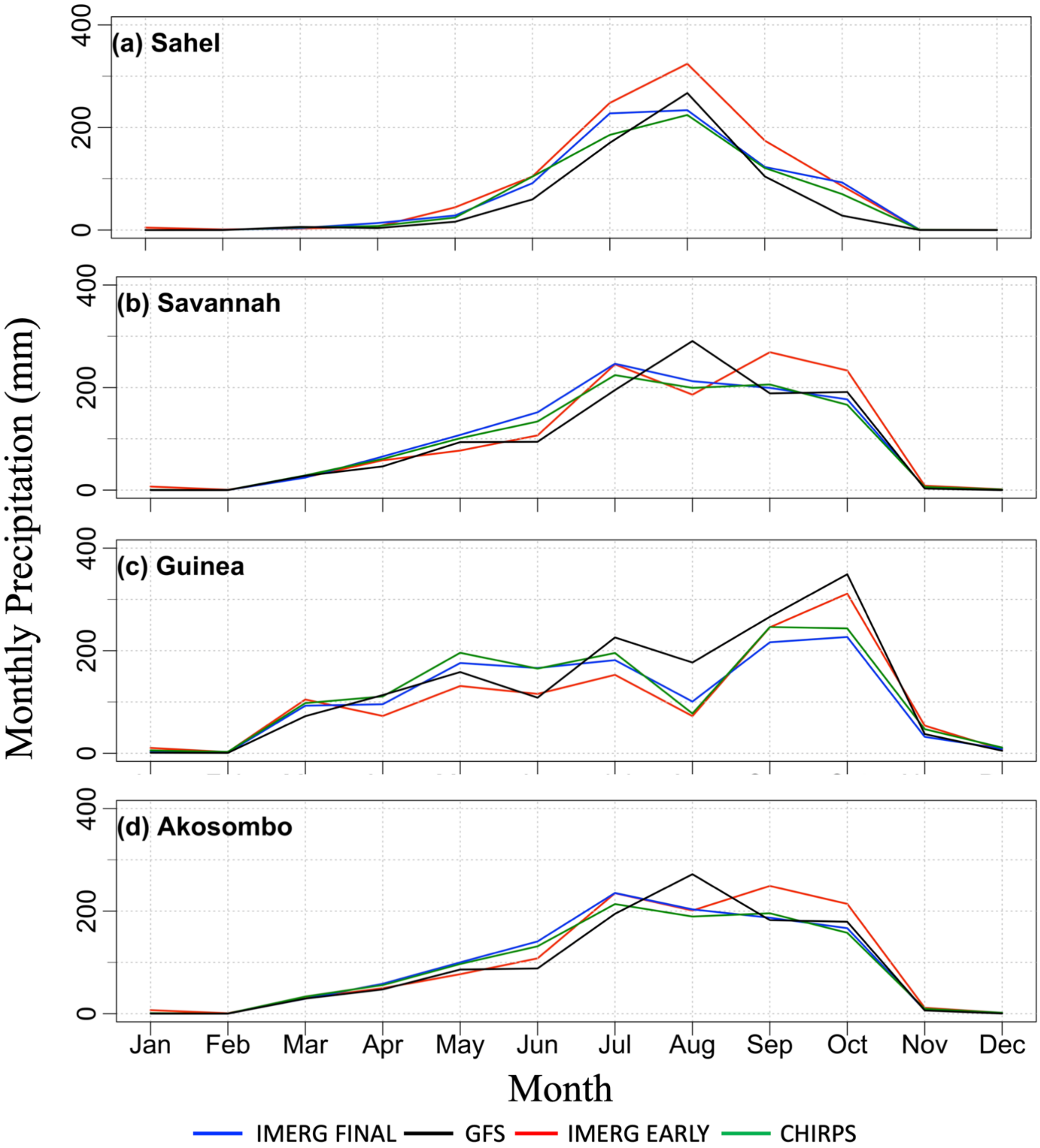


Figure 7.3. Monthly precipitation regime for the following regions of Volta River Basin: (a) Sahel climate, (b) Savannah climate, (c) Guinea Coastal climate, and (d) watershed of the Akosombo dam. Analyses are based on rainfall fields derived from IMERG Final, 1-day-lead GFS, IMERG Early, and CHIRPS. The time period covers from 15 June 2019 to 15 June 2020.

Averaging rainfall over the watershed of the Akosombo dam (i.e. averaging across all three climate zones) shows that the GFS forecasts capture well the seasonal rainfall pat-tern, but with some underestimation in the first few months of the rainy season, some overestimation in August, and almost perfect agreement from September until November. The GFS' overestimation bias in August was the result of overestimation bias in all climatological zones. The GFS' underestimation bias in the first few months of the rainy season was also the result of similar underestimation bias in al climatic zones. However, the GFS' almost perfect agreement with IMERG Final in the second half of the rainy season (September to November) comes as a result of the cancellation of the negative bias in the dry Sahel zone by the positive bias in the wet Guinea Coastal zone. In contrast to GFS, IMERG Early provide accurate results in the first half of the rainy season including August, but overestimates in the second half. Therefore, while IMERG Early outperforms GFS in the first half of the rainy season, GFS outperforms IMERG Early in the second half.

7.2.2 Dependence of Forecast Performance on Precipitation Rate

Figures 7.4a and 7.4b present the evaluation metrics of 1-day lead GFS using IMERG Final as reference, for the first half (March through July) and the second half (August through October) of the rainy season using basin-averaged values for Akosombo watershed. During March through July, GFS tends to underestimate by 21% overall. Moreover, the bias de-pends on the rain rate: the GFS is almost unbiased and close to the reference at low rain rates (< 5 mm/day), but has large underestimation bias for moderate and high rain rates. The overall RMSE is 113% of the mean rain rainfall rate. The linear correlation coefficient between GFS and IMERG Final is 0.59, indicating some agreement between the daily fluctuations of GFS and IMERG Final rainfall. The KGE is 0.48, which is on the border line between "poor" and “intermediate” skills for GFS in terms of its ability in capturing the basin-averaged daily rainfall values as expressed through the KGE statistics.

During the second half of the rainy season (August through October), GFS tends to overestimate by 14% overall, and moreover, the bias varies depending on the rain rate: the GFS has large overestimation bias at low rain rates (< 5 mm/day) and large underestimation bias at high rain rates (> 15 mm/day), but is unbiased at moderate rain rates (5 - 15 mm/day). The overall RMSE is 89% of the mean rainfall rate, the correlation between GFS and IMERG Final is 0.44, and the KGE of GFS is 0.44. Therefore, the GFS shows different bias characteristics during the first half and the second half of the rainy season.

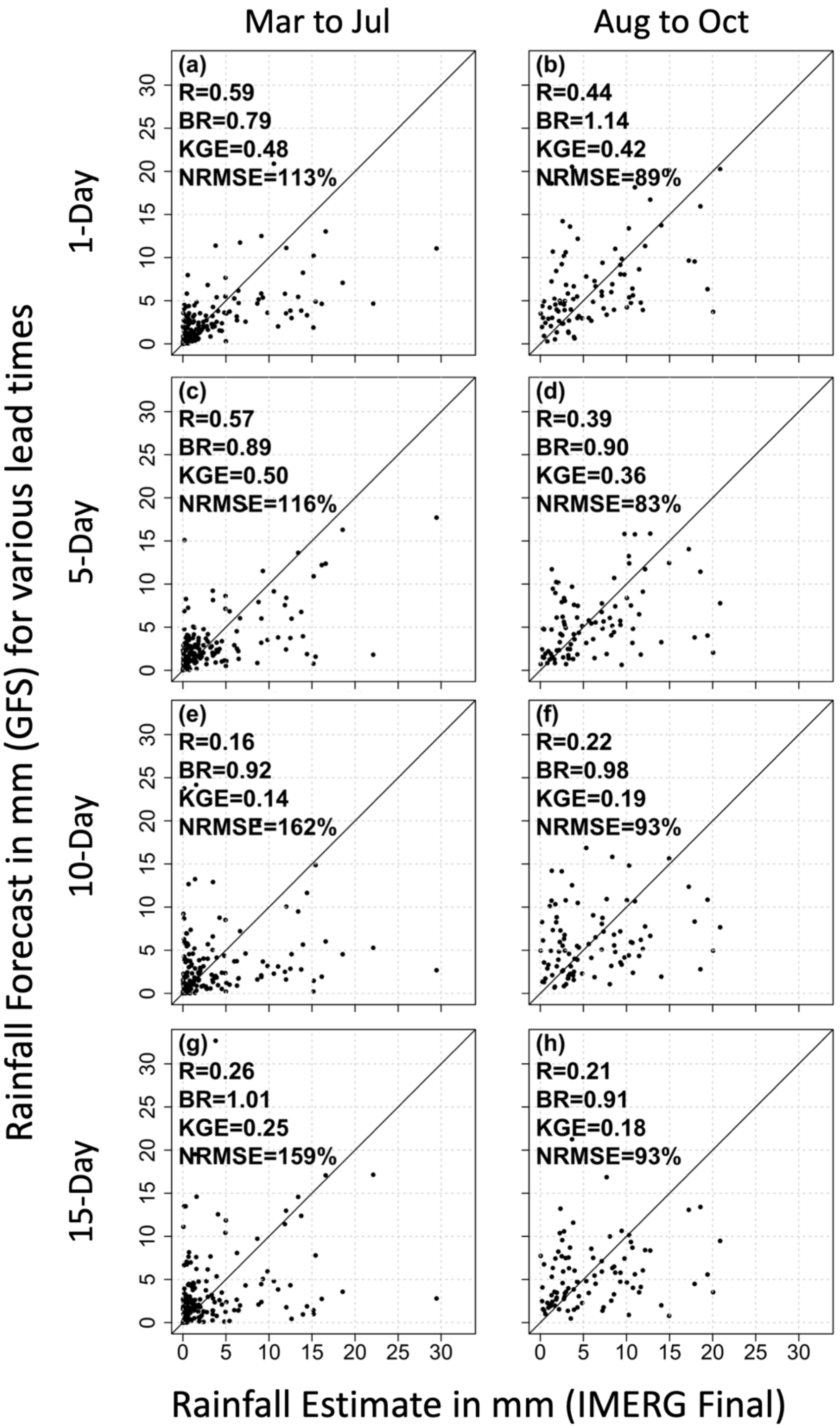


Figure 7.4. Scatterplot of sub-basin averaged accumulated precipitation forecast obtained from GFS forecast against corresponding values from IMERG Final, for various forecast lead times (1-day, 5-day, 10-day, and 15-day), and the first half (march through July) and the second half (August – October) rainy season.

7.2.3 What is the Effect of Lead Time on the GFS Forecast Performance?

Figures 7.4c through 7.4h present the evaluation metrics of daily rainfall forecast for 5-day, 10-day, and 15-day lead times, for the first half and second half of the rainy season. As the lead time increases, the forecast performance decreases. During March through Ju-ly, KGE decreased from 0.48 for 1-day lead, to 0.25 for 15-day lead; R decreased from 0.59 (1-day lead) to 0.26 (15-day lead); and NRMSE deteriorated from 113% (1-day lead) to 159% (15-day lead). Similarly, during August through October, KGE decreased from 0.42 (1-day lead), to 0.18 (15-day lead); R decreased from 0.44 (1-day lead) to 0.21 (15-day lead); and NRMSE deteriorated from 87% (1-day lead) to 93% (15-day lead). The reduction in performance from the 1-day lead to 5-day lead is very small, but is substantial at 10-day lead time forecast and beyond. The difference in performance between short and long-lead forecasts may be partly attributed to the data assimilation used in GFS (which improves initial conditions for short-lead forecasts) and lack of coupling with an ocean model as GFS uses only prescribed sea surface temperature (this would affect the performance of long-lead forecasts).

7.2.4 What is the Effect of Accumulation Timescale on the Forecast Performance?

Figure 7.5 presents the performance statistics of GFS rainfall forecast for 1-day, 5-day, 10-day, and 15-day accumulation timescales. For the March through July period, KGE increased from 0.48 for daily rainfall, to 0.74 for 15-day accumulated rainfall; R increased from 0.59 (daily) to 0.89 (15-day accumulation); and NRMSE improved from 113% (daily) to 32 (15-day accumulation). Similarly, for the August through October period, KGE in-creased from 0.42 (daily), to 0.88 (15-day accumulation); R increased from 0.44 (daily) to 0.89 (15-day accumulation); and NRMSE improved from 87 (daily) to 15 (15-day accumulation). Therefore, the GFS forecast performance substantially increases with the temporal aggregation scale. We point out that aggregation over longer temporal scales involves the use of data from both short-lead and long-lead forecasts, a situation that leads to two contradictory error natures. On the one hand, introducing data from long-lead times brings larger errors but, on the other hand, the temporal aggregation tends to average out the errors. This study shows that the effect of averaging out the errors outweighs the additional error introduced by the use of long-lead forecast. The rate of improvement in forecast accuracy with increasing lead time is bigger at short timescales than at long timescales. As the aggregation timescale increases from 1 day to 5 days, the NRMSE drops from 113% to 50%, and KGE jumps from 0.48 to 0.69. On the other hand, as aggregation timescale in-creases from 10 days to 15 days, NRMSE changes only slightly from 36% to 32% and KGE from 0.75 to 0.74.

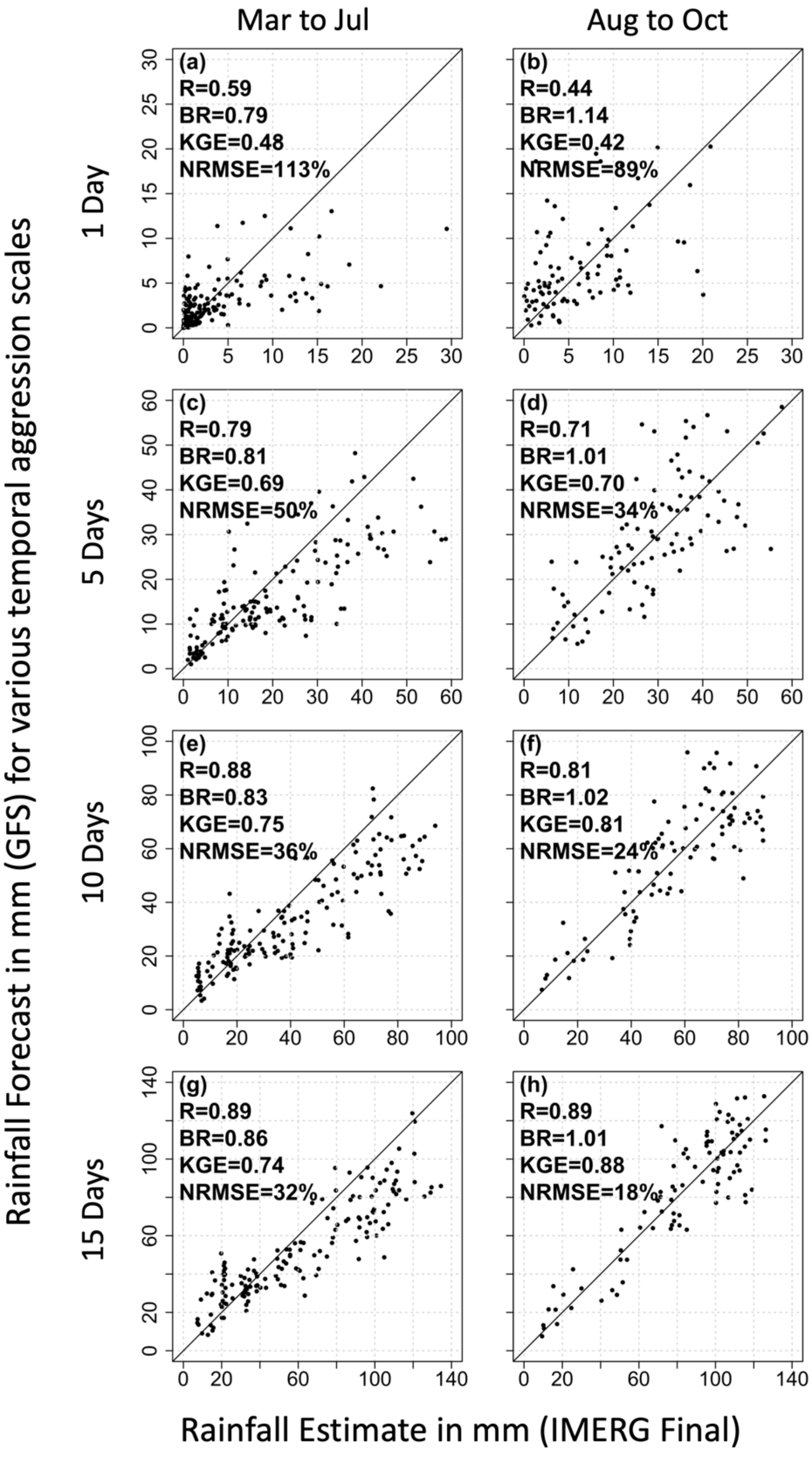


Figure 7.5. Scatterplot of sub-basin averaged accumulated precipitation forecast obtained from GFS forecast against corresponding values from IMERG Final, for various aggregation temporal scales (1 day, 5 days, 10 days, and 15 days), and the first half (March through July) and the second half (August – October) rainy season.

7.2.5 Comparison of the performance of IMERG Early and GFS

The performance statistics of GFS indicate the need for calibrating GFS forecasts in order to improve their accuracy. One way to achieve this could be through the use of post-processing techniques that involve rainfall estimates that have relatively better accuracy and are available in near-real time. Here, we assess the performance of IMERG Early, and compare it with the performance of GFS. Figure 7.6 provides the scatterplot of IMERG Early against IMERG Final, for 1-day, 5-day, 10-day, and 15-day aggregation temporal scales. As expected, due to the similarity of both algorithms, the correlation between the two estimates is very high, in excess of 0.90. However, IMERG Early tends to underestimate by about 20% in the first half of the rainy season, and overestimate by about 20% in the second half of the rainy season. The underestimation by IMERG Early in the first half of the rainy season is primarily due to the underestimation of medium rainfall rate. On the other hand, the overestimation of IMERG Early in the second half of the rainy season is due to its overestimation of high rain rate. Comparison of the performance statistics of GFS and IMERG Early indicates that the KGE statistics is higher than that for GFS, which is attributable to the high correlation (and variability ratio) of IMERG Early and IMERG Final due to the similarity of both algorithms. In terms of the bias statistics, IMERG Early outperforms GFS during the first half of the rainy season, but GFS outperforms IMERG Early during the second half. Therefore, IMERG Early has better capabilities in reproducing the temporal dynamics of rainfall, however, in terms of bias, it is better than GFS in some months but worse than GFS in some other months.

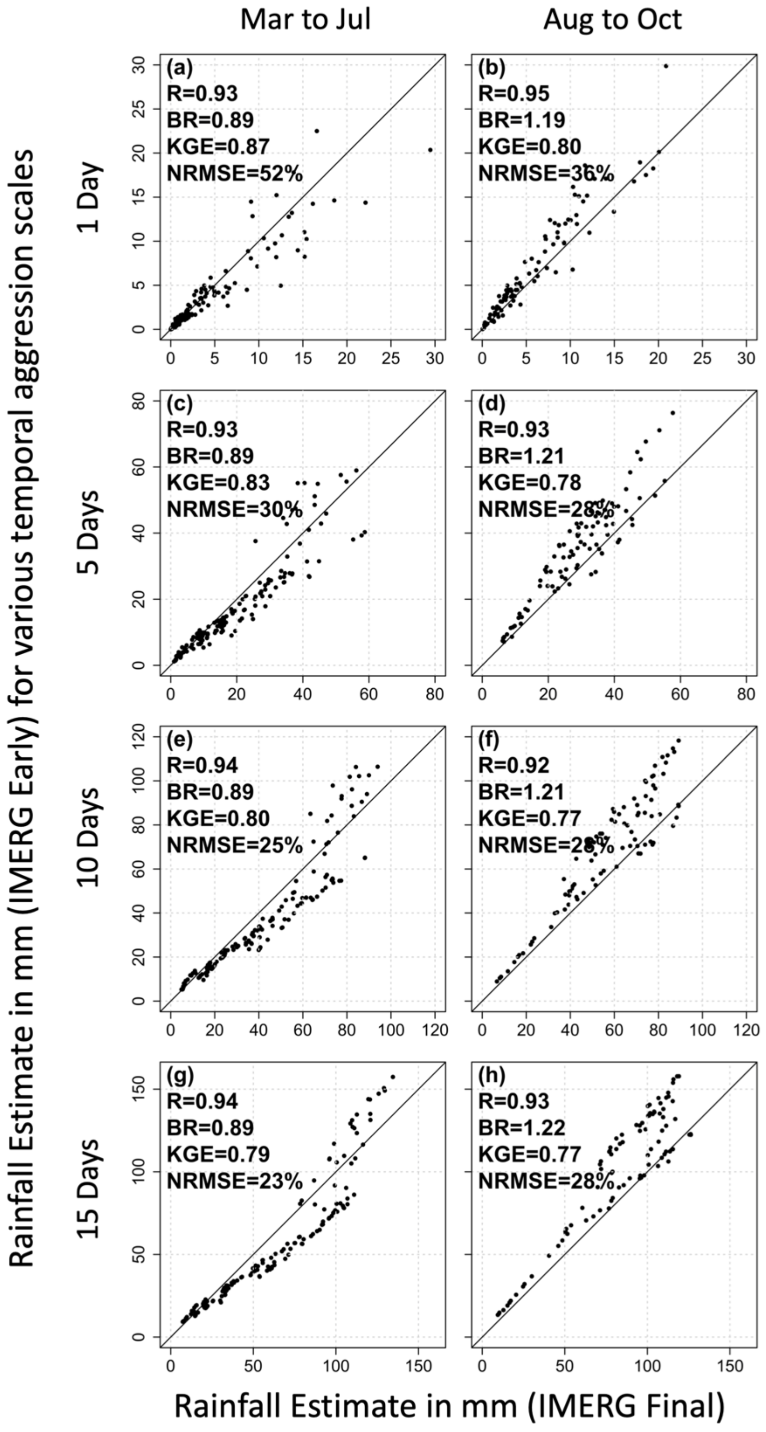


Figure 7.6. Same as in Figure 5 but for IMERG Early evaluation.

7.2.6 How is the Performance of GFS Affected if the Reference Product is Changed From IMERG Final to CHIRPS?

We acknowledge that the reference dataset used in our evaluation (i.e., IMERG Final) has its own estimation errors. We conducted additional assessment to evaluate the performance of GFS using CHIRPS rainfall products as reference. Table 7.1 shows the performance statistics of GFS for different lead times, using IMERG FINAL and CHIRPS, separately, rainfall products as reference. The pattern of GFS performance with respect to lead time is similar when either rainfall product is used as a reference. The correlation and KGE statistics of GFS obtained using IMERG Final as reference are very close to those statistics of GFS obtained using CHIRPS as reference. In terms of bias and NRMSE statistics, the performance of GFS computed using the two products as reference have some differences in actual magnitudes. The NRMSE statistics computed using CHIRPS as reference are lower than those obtained using IMERG Final as reference. The bias statistics computed using CHIRPS as reference could be better or worse than those obtained using IMERG Final as reference, depending on the month. Therefore, our results show that the overall performance of GFS remains the same if the reference product were to be changed from IMERG Final to CHIRPS, indicating the robustness of IMERG Final as reference product.

Table 7.1 Performance statistics of GFS forecast for various lead times (1-day, 5-day, 10-day, and 15-day) using IMERG Final (CHIRPS) rainfall products as reference, in terms of correlation, bias ratio, and NRMSE.

Lead time of GFS forecast	**Correlation**	**Bias Ratio**	**KGE**	**NRMSE (%)**
Time Period: March through July				
1-day	0.59 (0.57)	0.79 (0.65)	0.48 (0.44)	113 (87)
5-day	0.57 (0.50)	0.89 (0.73)	0.50 (0.42)	116 (91)
10-day	0.16 (0.17)	0.92 (0.76)	0.14 (0.11)	162 (121)
15-day	0.26 (0.30)	1.01 (0.83)	0.25 (0.22)	159 (118)
Time Period: August through October				
1-day	0.44 (0.50)	1.14 (1.00)	0.42 (0.49)	89 (71)
5-day	0.39 (0.39)	0.90 (0.79)	0.36 (0.36)	83 (72)
10-day	0.22 (0.18)	0.98 (0.86)	0.19 (0.17)	93 (81)
15-day	0.21 (0.21)	0.91 (0.80)	0.18 (0.19)	93 (80)

7.3 Conclusion

This study has evaluated the accuracy of medium-range (1-day to 15-day lead time) forecasts available from the Global Forecast System (GFS), for the watershed of Akosombo dam in the Volta basin, west Africa, using satellite-gauge merged IMERG Final and CHIRPS rainfall products. Additionally, the performance of GFS is compared to the per-formance of the satellite-only IMERG Early product to see if the latter can potentially be used in the post-processing of GFS.

The performance statistics of 1-day lead watershed-averaged GFS forecast are: KGE = 0.42 to 0.49, correlation = 0.44 to 0.59, NRMSE = 71% to 114%. The performance of GFS depends on climate zone, month, rain intensity, lead time, and temporal aggregation scale. The GFS tends to underestimate (by 20%) in the dry Sahel region, overestimate (by 16%) in the Guinea Coastal region, while producing a relatively more accurate result in the Savannah region. Averaging rainfall over the watershed of the Akosombo dam (i.e. averaging across all three climate zones) shows that the GFS forecasts capture well the seasonal rainfall pattern, but tend to underestimate rainfall (by 21%) in the first half of the rainy season (March through July) and overestimate rainfall (by 14%) in the second half of the rainy season (August through October). In the first half the rainy season, the GFS is unbiased at low rain rates, but has large underestimation bias for moderate and high rain rates. In the second half of the rainy season, the GFS has overestimation bias at low rain rates and large underestimation bias at high rain rates, but is unbiased at moderate rain rates. The GFS forecast performance decreases with increasing lead time. The reduction in forecast performance from the 1-day lead to 5-day lead is very small, but is substantial at 10-day lead time forecast and beyond. The GFS forecast performance substantially increases with the temporal aggregation scale. The rate of improvement in forecast accuracy with in-creasing lead time is bigger at short timescales than at long timescales. The 15-day total GFS forecast has the following performance statistics: KGE = 0.74 to 0.88, correlation = 0.89, and bias ration = 0.86 and 1.01. Overall, we conclude that the GFS forecasts, at 15-day accumulation timescale, have acceptable performance, although they tend to underestimate moderate and high rain rates during March to July. IMERG Early products can be considered for use in post-processing of GFS products to improve the bias during March to July.

CHAPTER 8

Evaluation of Medium-Range Forecasts (GFS) over Congo

8.1 Study Region

The Congo River Basin (Figure8.1), located in the equatorial region of Africa, is the second larger river system in the world, both in terms of drainage area (~3.7 106 km2) and discharge (mean annual flow of ~ 40,500 m3/sec) (Laraque et al. 2009, 2013). The basin is one of the three main convective centers in the Tropics and receives an average annual rainfall of about 1,500 mm/yr (Hastenrath 1985). The basin is also the second largest rain forest area in the world (after Amazon) with its 45% of the land area covered with forecasts (Verhegghen et al. 2012; Lossow 2017). The Congo Basin is a transboundary river basin that encompasses the following riparian countries: Democratic Republic of the Congo (accounting for 62% of the basin's area), Central African Republic (11%), Angola (8%), the Republic of the Congo (7 %), Zambia (5 %), Tanzania (4 %), Cameroon (2 %), and Gabon, Burundi and Rwanda (< 1%) (Wohl, 2007).

The main seasonal peak in precipitation falls around December-March in the basin areas well south of the equator, and around July-October in the regions north of the equator (Mahe et al. 2013). The north-south movement of rainfall has been assigned to the seasonal migration of the Intertropical Convergence Zone (ITCZ) and mesoscale convective systems bounded by the seasonally migrating Northern and Southern Africa Easterly Jets and enhanced by orography, i.e., the Bowel like topography of the Congo Basin (Jackson et al. 2009). The high flow rate and steep gradients created enormous opportunities for the Congo River and its tributaries to generate hydropower, although to date, less than 3% of this potential has been tapped. For this study, we selected four dams, all hydropower, that represent different parts of the basin (Fig. 1, Table 1): Inga I dam in Democratic Republic of the Congo (DRC) (20 Mm3 storage capacity, and generates 58.5 MW power), Djoue Dam in Republic of the Congo (8.5 Mm3, 5 MW), Ruzizi I dam in DRC (1.5 Mm3, 27 MW), and Koni dam in DRC (28 Mm3, 42 MW).

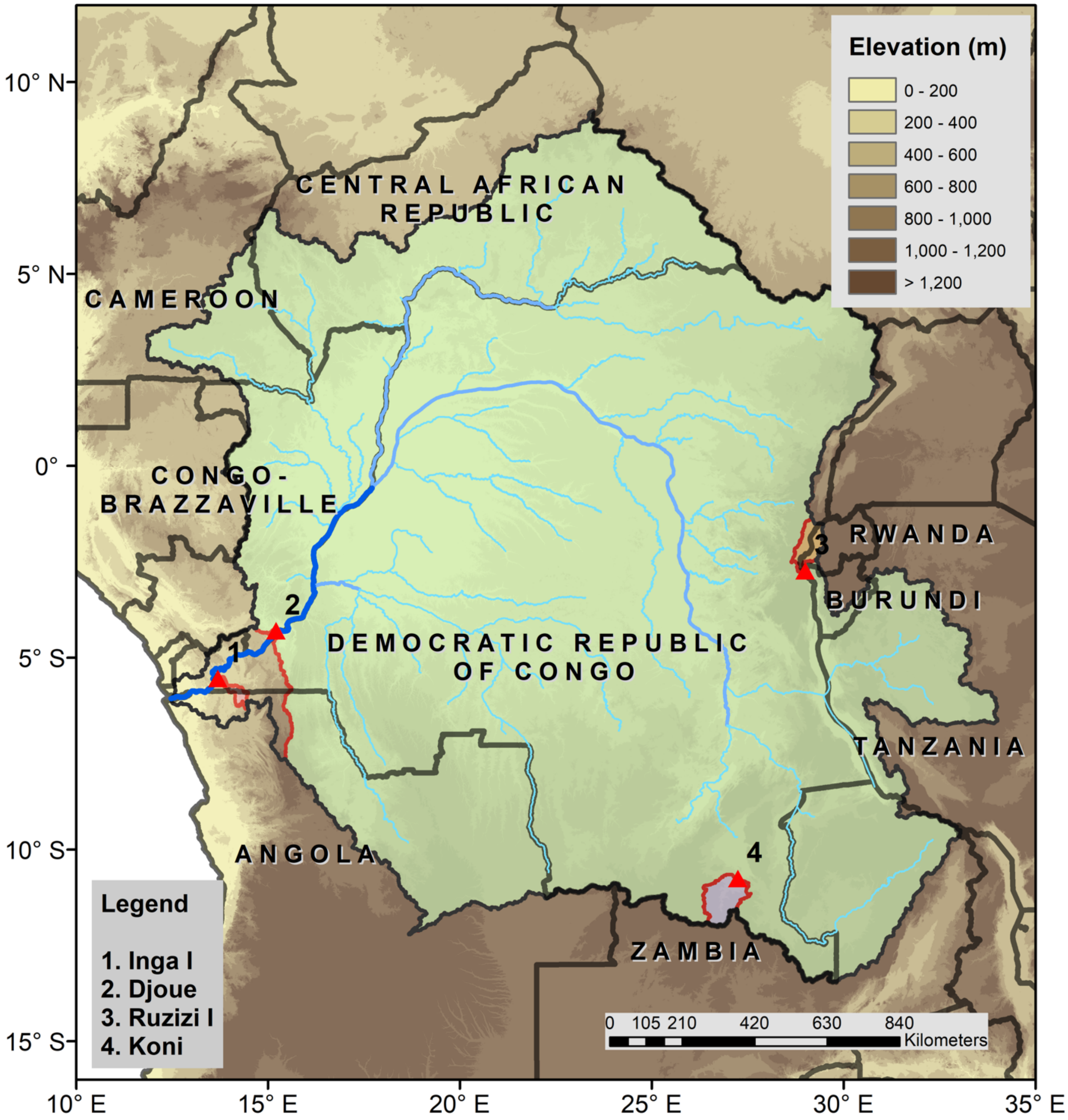


Figure 8.1 Map of Congo River Basin, with the location of selected dams/reservoirs: (1) Inga I, (2) Djoue, (3) Ruzizi I, and (4) Koni, and the drainage basins defined by the dam locations (see the colored areas between dams).

Table 8.1. Selected dams and their watershed characteristics

Dams	Country	Operational Since*	Capacity (million m^3)*	Power (MW)*	Area of Drainage Basins (km^2)**	Elevation of Drainage Basin (m)**
Inga I	Democratic Republic of the Congo	1972	20	58.5	2566	443
Djoue	Congo	1956	8.5	5	3605001	732
Ruzizi I	Democratic Republic of the Congo	1959	1.46	27.3	7833	1776
Koni	Democratic Republic of the Congo	1949	28	42.12	12604	1287

* information obtained from the Global Reservoir and Dam Database (Lehner et al. 2011) and Food and Agriculture Organization of the United Nations (FAO)'s Global Information System on Water and Agriculture (AQUASTAT).
** Calculated from HydroSEHDS (Lehner et al. 2008).

8.2 Results and Discussion

8.2.1 Monthly Cycle of Rainfall

Figure 8.2 presents the monthly time series for watershed-averaged rainfall derived from the 1-day lead GFS forecast and satellite products (IMERG Final, IMERG Early, and CHIRPS), for each of the dams. For the Inga I dam, whose watershed is located in the western part of the watershed and around 5°S latitude, IMERG Final shows rainy season from October through April, with dry period of four months (May through September), and peak rainfall slightly under 300 mm. The GFS over Inga I shows similar pattern, but tends to overestimate monthly rainfall, with large overestimations in October through December. For the Ruzizi I dam, whose watershed is located in the eastern part and relatively close to the Equator (2.5°S), the IMERG Final reference product shows a similar seasonal pattern as for Inga I. The GFS over Ruzizi I gives very large overestimation bias with forecasts as much as twice the reference, for a number of months. For the Koni dam, whose watershed sits in the southern east around 14°S, IMERG Final shows rainfall season from November through March, with a peak rainfall of around 250 mm in December, and extended dry period, from April until October. The GFS over Koni gives very large overestimation bias (with peaks reaching 500 mm/month from December through March). For the Djoue watershed, which covers a very large watershed (with drainage area of 3.5 million km2), IMERG Final shows bimodal rainfall pattern covering most of the year, and relatively short low-rainfall (but still wet) period. The GFS over Djoue captures the time series with some overestimation bias throughout the year. Overall, the GFS forecasts show overestimation over all dam watersheds, but the overestimations are larger at the Ruzizi I and Koni watersheds, located in the eastern and southeastern parts of the watershed.

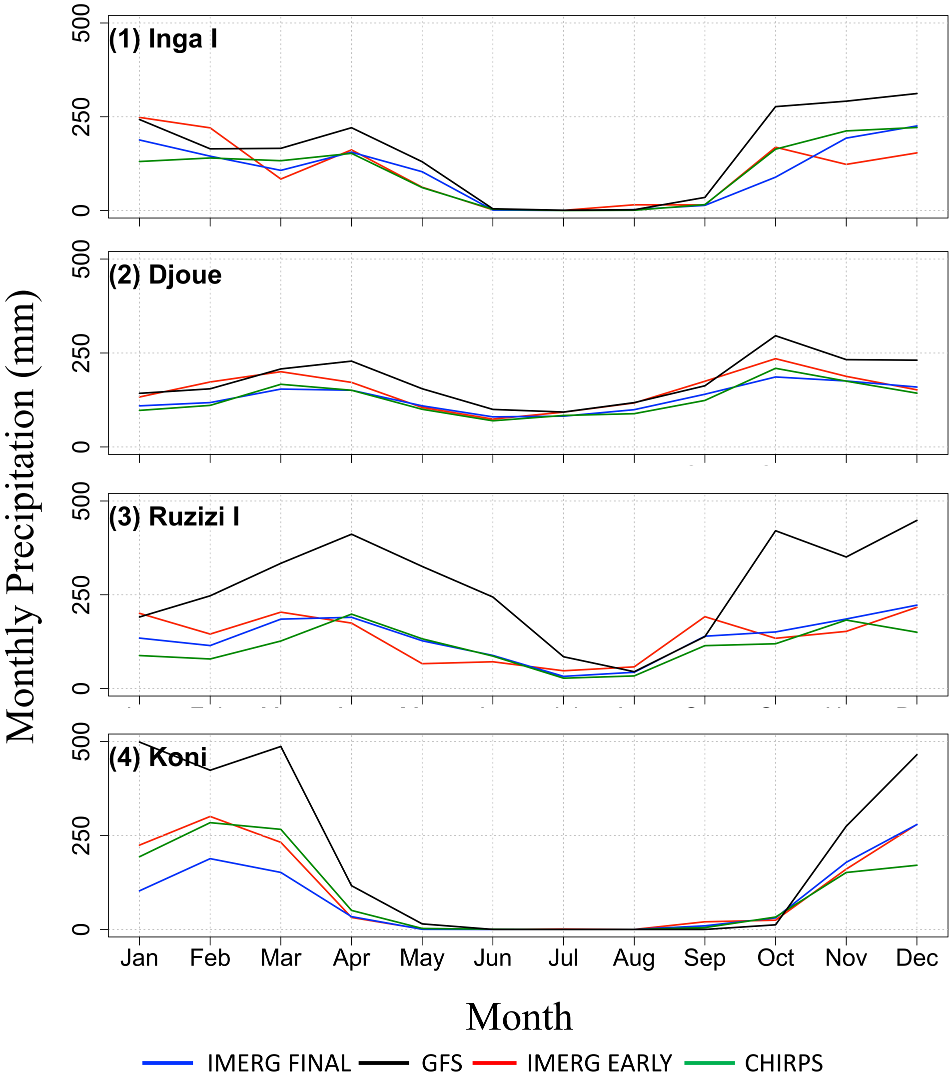


Figure 8.2. Monthly time series of sub-basin averaged precipitation (mm), for the period 15 June 2019 to 15 June 2020, derived from IMERG Final, GFS (1-day lead time), IMERG Early, and CHIRPS, for watersheds of four dams: (1) Inga I, (2) Djoue, (3) Ruzizi I, and (4) Koni.

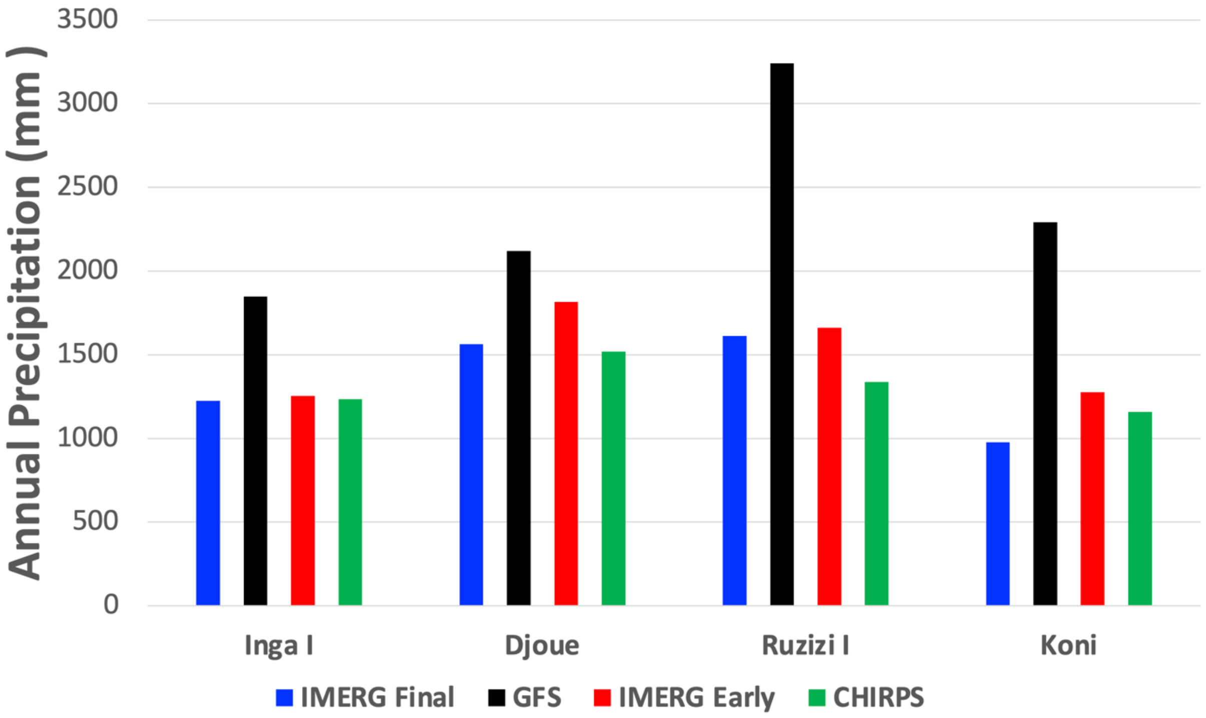


Figure 8.3. Sub-basin averaged annual precipitation (mm) for the period, 15 June 2019 to 15 June 2020, for each of the Congo's sub-basins, derived from the 1-day lead GFS forecast and different satellite precipitation products.

8.2.2 Annual Rainfall

Here, we aggregate the 1-day lead GFS forecasts to annual time scale and compare the results against corresponding annual precipitation estimates from IMERG Final (Figure 8.3). According to IMERG Final, the annual watershed-average rainfall is 1224 mm (Inga I), 1564 mm (Djoue), 975 mm (Koni), and 1613 mm (Ruziz I). The 1-day lead GFS forecasts overestimate the annual rainfall in all cases, with bias ratios ranging from 36% (Djoue), to 51% (Inga I), to 134% (Koni, and to 101% (Ruziz I). The overestimation is relatively small for the largest watershed (Djoue) and the watershed located in the western part (Inga I), while it is larger for the small watersheds located in the eastern (Ruzizi I) and southern (Koni) parts of the watershed.

8.2.3 Daily Time Series of Rainfall

Figure 8.4 presents the daily time series of watershed-averaged rainfall. According to IMERG Final, the daily rainfall pattern over Inga I has a number of large spikes (well above 300 mm/day), but the GFS misses the majority of these large spikes and gives relatively smooth fluctuation. The IMERG Final-derived daily rainfall pattern over Ruzzi I dam watershed is as also spiky - the corresponding GFS forecast also gives spiky daily rainfall pattern but there is a mismatch in the spike occurrences between GFS and IMERG Final. For the Koni watershed, IMERG Final shows a few spikes in daily rainfall - the corresponding GFS shows large substantial overestimation in most of the cases. For the Djoue watershed, IMERG Final shows low values of daily rainfall, with no pronounced spikes - the corresponding GFS forecasts capture well this pattern.

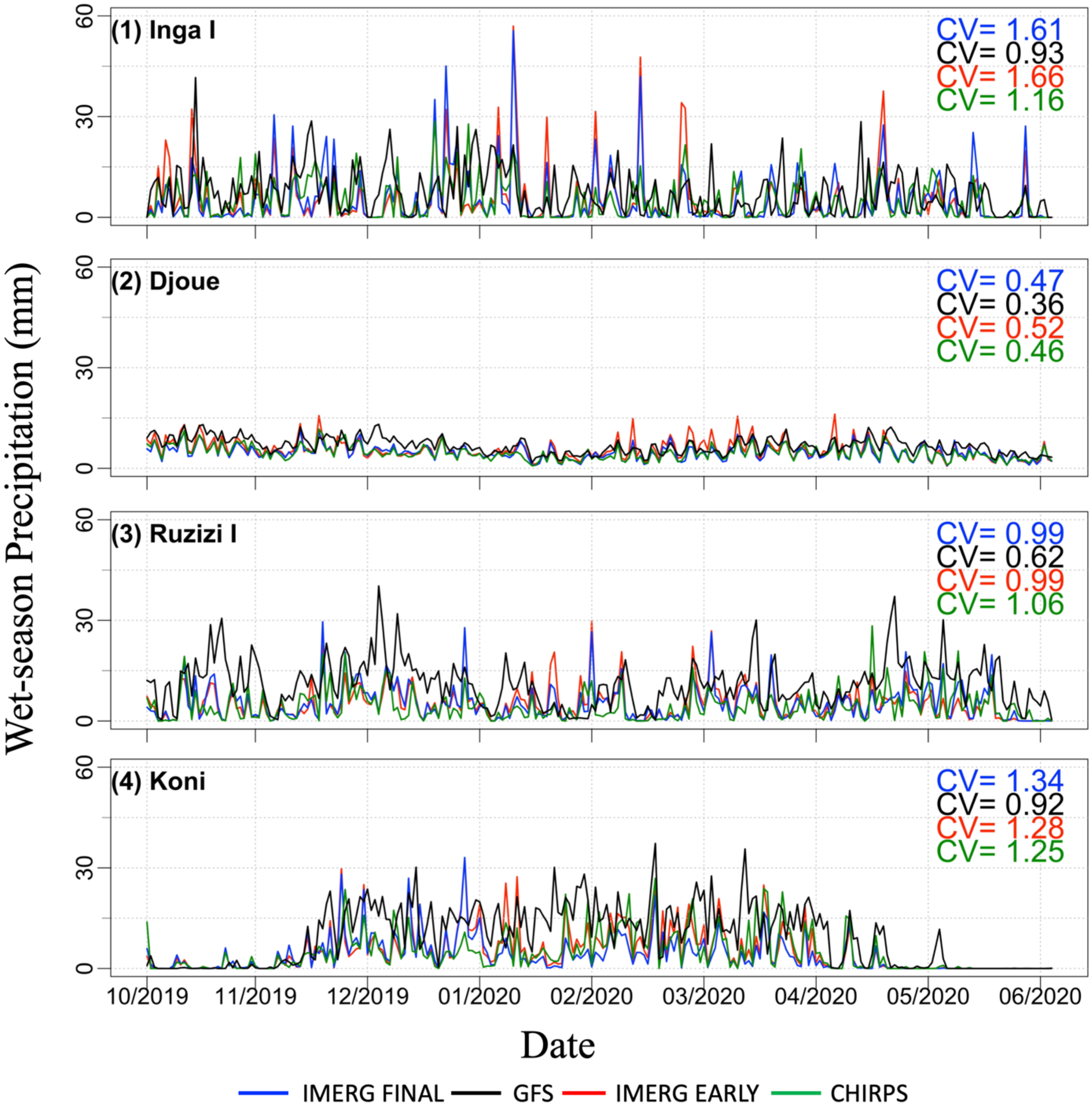


Figure 8.4. Time series of sub-basin averaged precipitation total (mm) for the wet period (October – May), for each of the dam watersheds, as derived from various precipitation products. The Figure also shows the coefficient of variation (CV) as a measure of temporal variation.

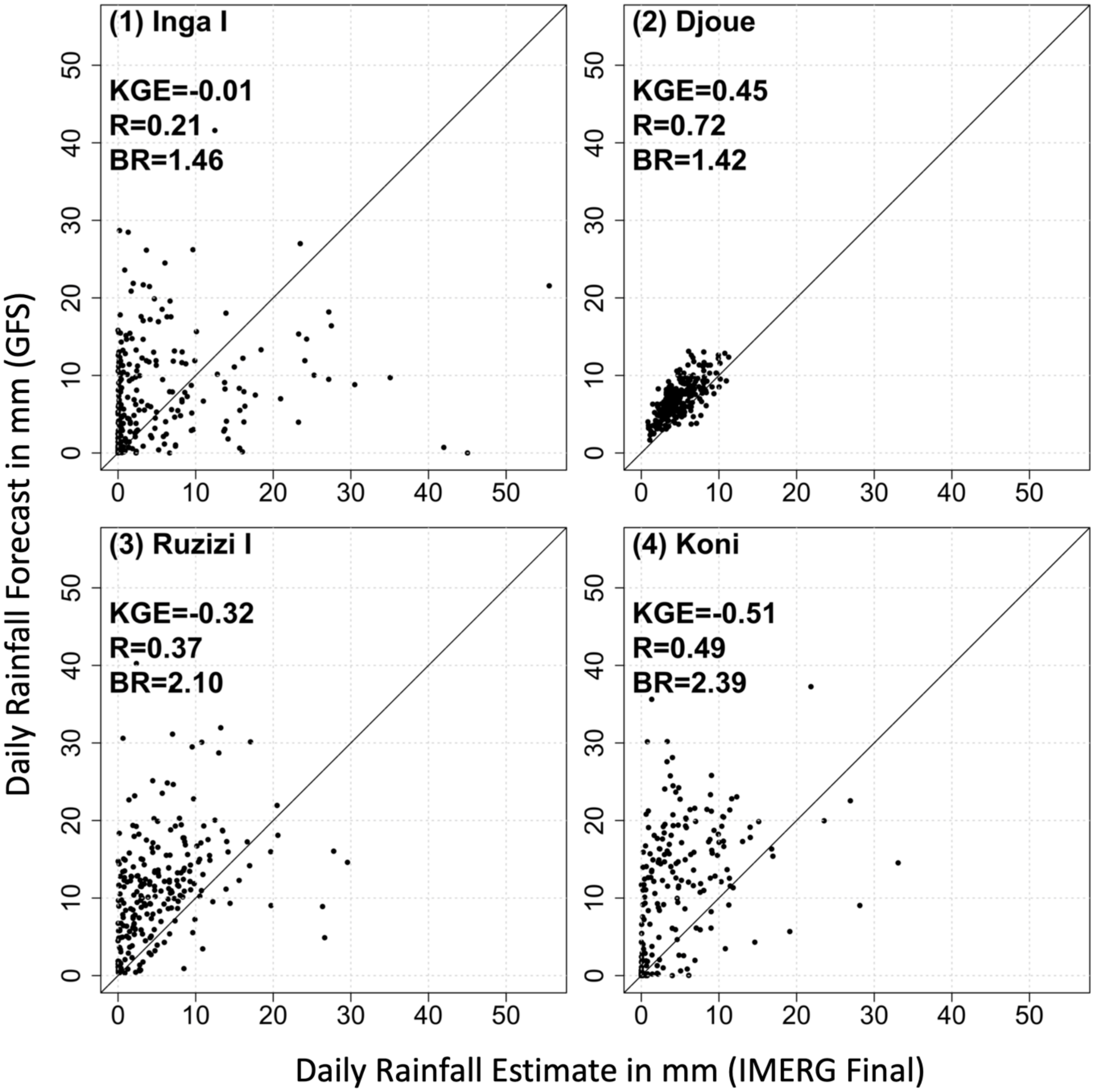


Figure 8.5. Scatterplot of watershed-averaged daily precipitation forecast obtained from the 1-day lead GFS forecast against corresponding values from IMERG Final, for four dam watersheds.

8.2.4 Dependence of Forecast Performance on Precipitation Rate

Figure 8.5 presents the scatterplot of 1-day lead GFS forecasts and IMERG Final at daily and watershed-average scales. Over the Inga I dam watershed, the GFS underestimates the large rainfall events (above 25 mm/day) gives large false alarms, and has large scatter around light to moderate rainfall, resulting a correlation of only 0.21 and KGE of -0.01, indicating that the GFS has very poor skill in capturing daily watershed-averaged rainfall. Over the Ruzizi I dam watershed, the GFS has substantial overestimation for almost the majority of the daily rainfall, with correlation coefficient of 0.37, resulting in KGE < 0, indicating very poor skill in capturing daily watershed-averaged rainfall. Similarly, over the Koni watershed, the GFS shows poor agreement with IMERG Final, with large false alarms, overestimation bias in almost all cases, and overall large negative KGE, indicating very poor skill. Over the Djoue watershed, the GFS shows relatively good agreement with the IMERG Final, with a KGE of 0.45, but with overestimation by about 40%.

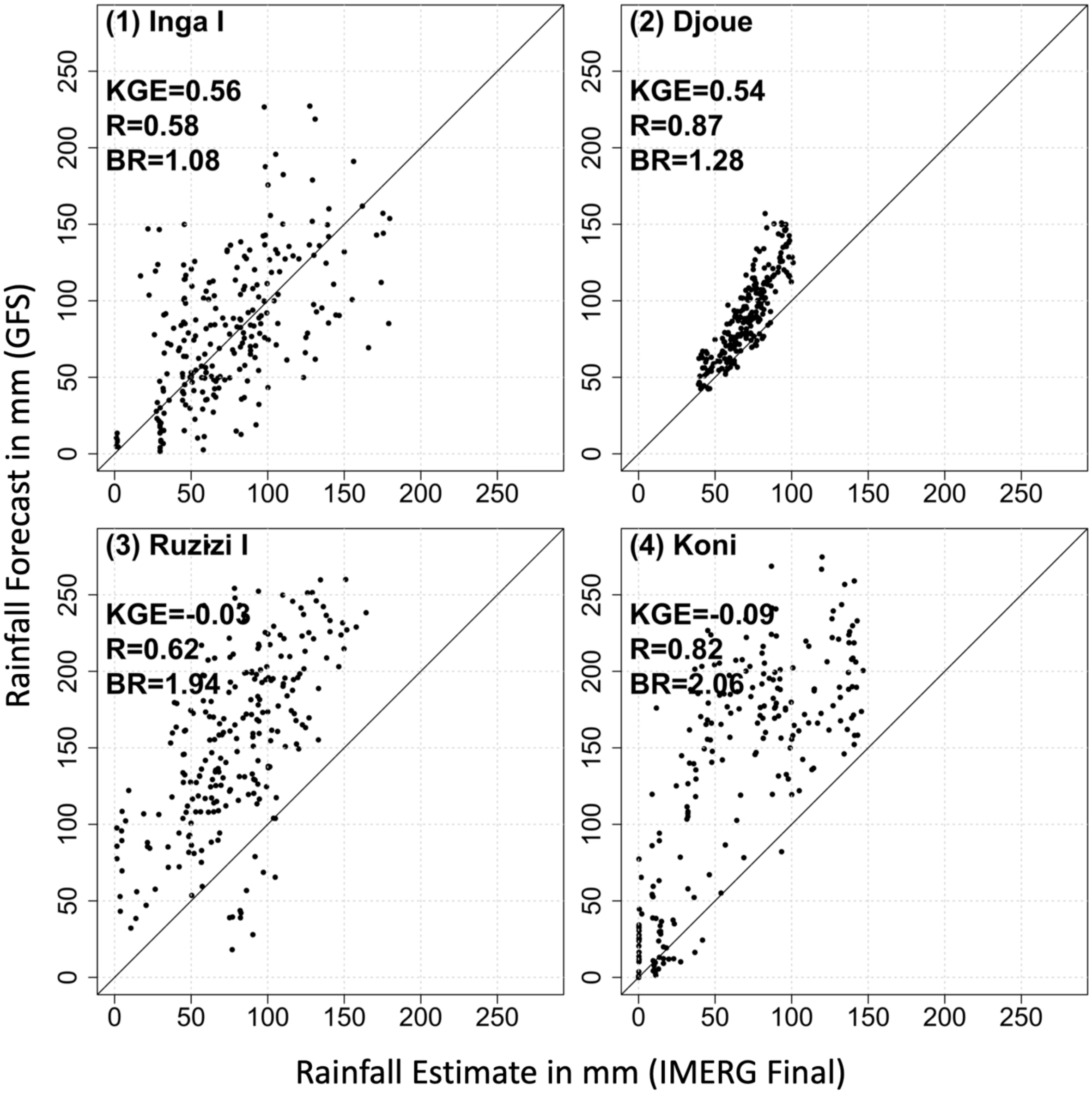


Figure 8.6. Scatterplot of watershed-averaged daily precipitation forecast obtained from the 1-day lead GFS forecast against corresponding values from IMERG Final, for four dam watersheds.

8.2.5 Effect of Temporal Aggregation Scale on Forecast Performance

Figure 8.6 presents the scatterplot of watershed-averaged 15-day total rainfall obtained from the GFS forecasts against IMERG Final estimates, for each dam watershed. The 15-day total forecast is obtained by adding multiple lead-time daily forecasts: 1-day lead, 2-day lead, and all the way up to 15-day lead. Over the Inga I dam watershed, the 15-day total GFS forecast has better correlation (R = 0.51 compared to R = 0.18 for daily rainfall), better bias (BR = 1.08 compared to 1.46), resulting in better KGE (0.56 compared to -0.01), indicating "intermediate" skill in capturing the 15-day total rainfall as reported by IMERG Final. Over the Ruzizi I watershed, the 15-day total GFS forecast has better performance statistics compared to daily, but still the resulting KGE (-0.03) is about borderline between very poor and poor skill. Over the Koni watershed, the 15-day total GFS shows some improvement (yet unsatisfactory) over the daily forecast performance, with still negative KGE and very high overestimation bias for all rain rates. Over the Djour watershed, the 15-day total GFS shows improvement in performance (KGE = 0.54 compared to 0.45 at daily timescale), resulting in “intermediate” skill for GFS forecasts.

8.2.6 Effect of Lead Time on Forecast Performance

In Figure 8.7, we show the Kling-Gupta Efficiency (KGE) and its components for 5-day total rainfall forecast of GFS for three different lead time periods, 1-5 day, 6-10 day, and 10-15 day. The KGE values at 1-5 day lead time are almost identical to those at 11-15 day lead time. The KGE values at 6-10 day lead time are slightly higher. The breakdown of the KGE scores (BR, R and γ) reveals that the BR and γ statistics remain the same across all lead times. However, the correlation R statistic decreases with increasing lead time.

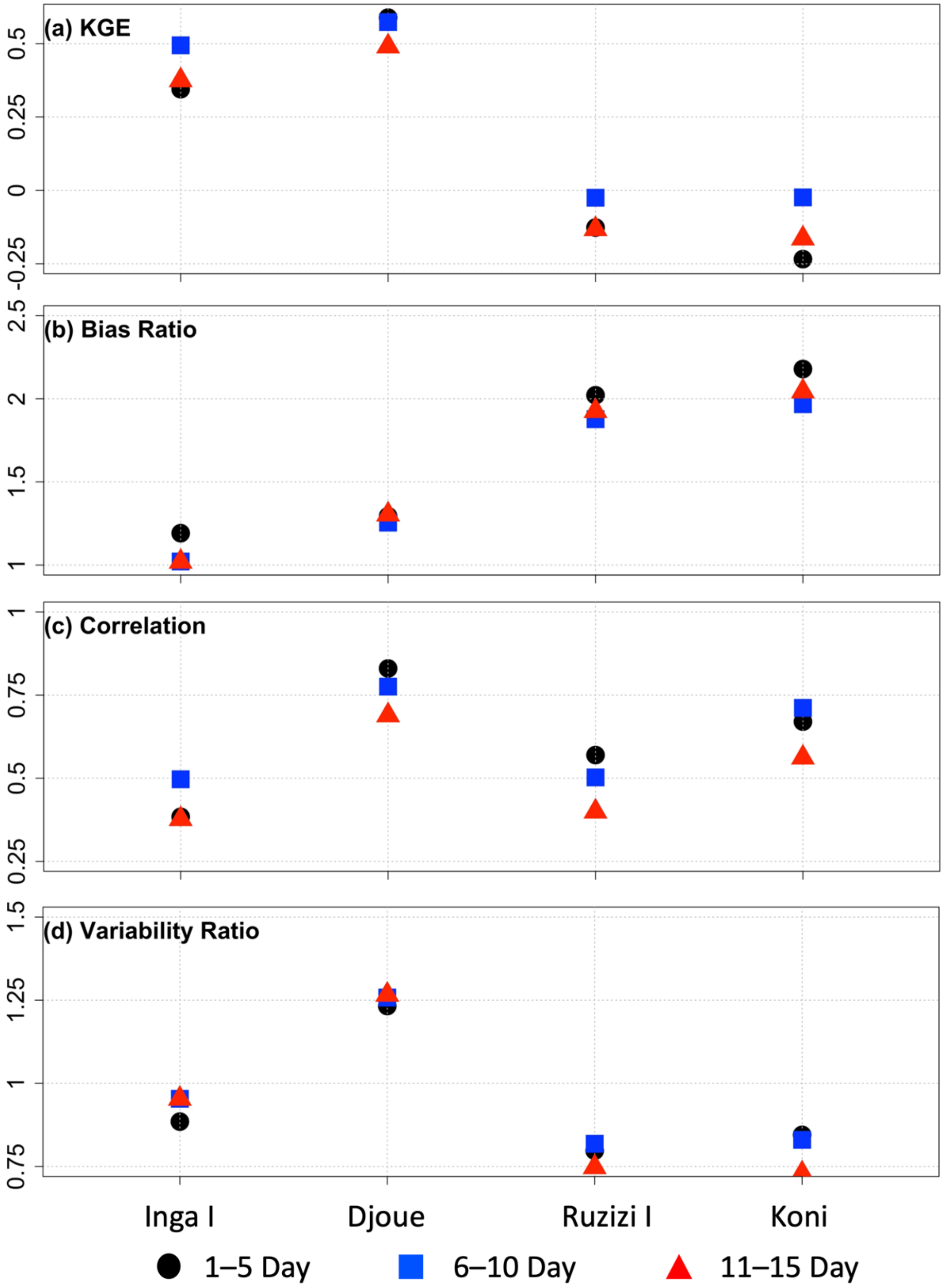


Figure 8.7. Kling-Gupta Efficiency (KGE) and its components for 5-day total rainfall forecast of GFS for three different lead time periods, 1-5 day, 5-10 day, and 10-15 day, for each dam watershed.

8.2.7 Comparison of the Performances of IMERG Early and GFS

The low performance statistics of GFS (e.g., KGE < 0 at three watersheds and KGE of 0.45 at Djoue, for 1-day lead) suggest the need for exploring ways of improving the GFS forecast accuracy. One such method is post-processing of GFS forecasts using rainfall estimates that have relatively better accuracy and are available in near-real time. Satellite-only products, such as IMERG Early, are available in near-real time. In this section, we explore how the performance of IMERG Early compare with the performance of GFS. The monthly timeseries of IMERG Early is more accurate than GFS (Figure 8.2). The annual values also indicate much better performance for IMERG Early than for GFS (Figure 8.3). IMERG Early also better capability in capturing the daily time series of rainfall than for GFS (Figure 8.4). The KGE values for IMERG Early (0.69 to 0.88) are higher than those for GFS (-0.51 to 0.45) in all dam watersheds (Figure 8.8). All the three components of KGE (bias, correlation, and variability ratio) are also higher for IMERG Early than for GFS (Figure 8.8). Therefore, all performance measures indicate that the IMERG Early has better performance than the GFS, which implies that the IMERG Early products could be considered for post-processing of GFS in order to improve the accuracy of GFS forecasts.

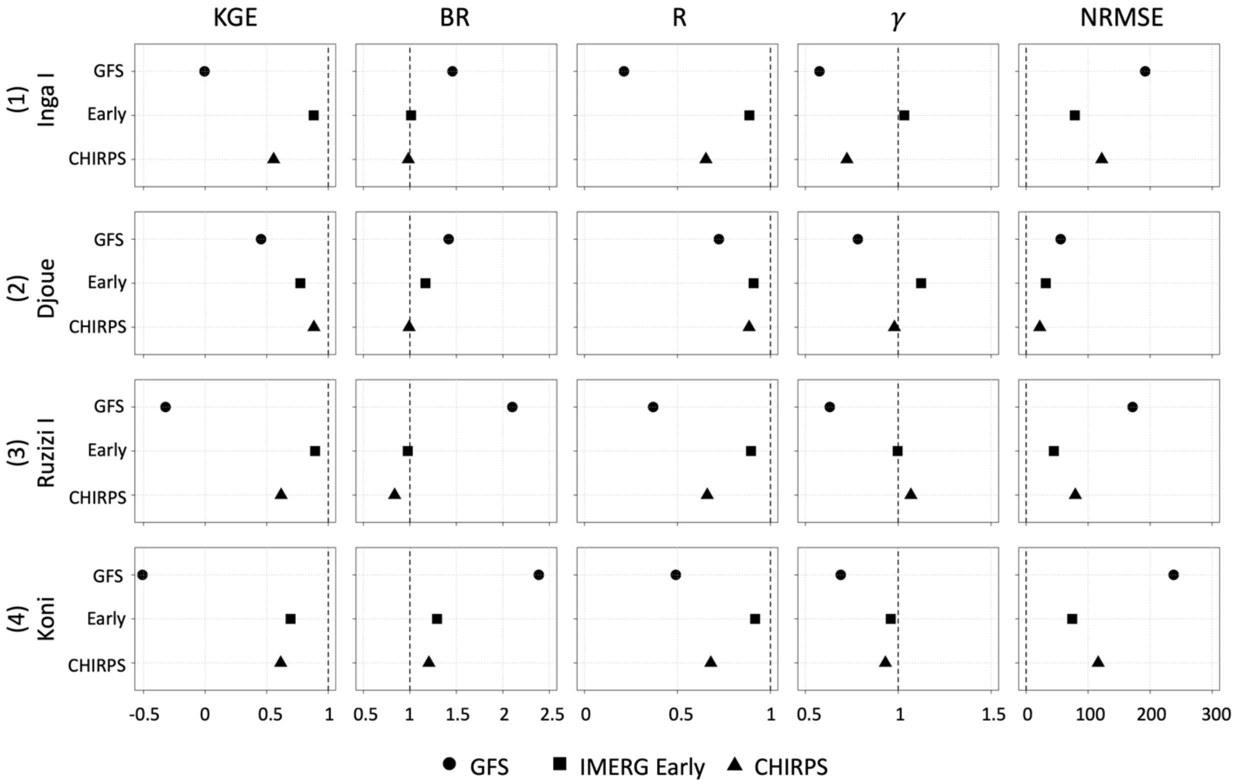


Figure 8.8. Summary of performance statistics (Kling-Gupta Efficiency KGE, Bias Ratio BR, correlation R, variability ratio γ, and root mean square error normalized by reference precipitation mean NRMSE) of 1-day lead GFS forecasts and different satellite products, during the wet periods, at different watersheds.

Table 8.2 Performance Statistics of 1-day Lead Daily GFS Forecast using IMERG Final (CHIRPS) Rainfall Products as Reference, in terms of Correlation, Bias ratio, and NRMSE.

Dam	Correlation	Bias Ratio	KGE	NRMSE (%)
Inga I	0.21 (0.33)	1.46 (1.48)	-0.01 (0.15)	192 (155)
Djoue	0.72 (0.72)	1.42 (1.43)	0.45 (0.45)	56 (57)
Ruzizi I	0.37 (0.30)	2.10 (2.51)	-0.32 (-0.72)	172 (221)
Koni	0.49 (0.57)	2.39 (1.98)	-0.51 (-0.10)	238 (181)

8.2.8 Performance of GFS if the Reference Product is Changed from IMERG Final to CHIRPS

We acknowledge that the reference dataset used in our evaluation (i.e., IMERG Final) has its own estimation errors. We conducted additional assessment to evaluate the performance of GFS using CHIRPS rainfall products as reference. Table 2 shows the performance statistics of GFS for different lead times, using IMERG Final and CHIRPS, separately, rainfall products as reference. The overall magnitude of GFS performance (as well as the variability of the performance across watersheds) is similar when either rainfall product is used as a reference. Both reference products reveal that the GFS forecast has very low KGE in the small three watersheds, and relatively higher KGE in the largest watersheds (Djoue). In addition to KGE, the KGE breakdown components remain about the same regardless of the reference product used. Therefore, our results show that the overall performance of GFS remains the same if the reference product were to be changed from IMERG Final to CHIRPS, indicating the robustness of IMERG Final as reference product.

8.3 Conclusion

This study has evaluated the accuracy of medium-range (1-day to 15-day lead time) forecasts available from the Global Forecast System (GFS), for the watershed of four major dams in the Congo basin, using satellite-gauge merged IMERG Final and CHIRPS rainfall products. Three of watersheds are small in size (2,570 to 12,600 km2), but are located spread in the eastern (Ruzizi I), western (Inga I), and southern (Koni) parts of the watershed. The 4th watershed (Djoue) is very large (3.6 million km2).

The 1-day lead GFS shows large overestimation bias in almost all cases, except for Inga I (where it still overestimates half of the cases). The overestimation is particularly very high (more than twice of the monthly as well as annual rainfall values) for the Ruzizi I and Koni watersheds, located in the eastern and southern parts of the watershed. The GFS shows large number of false alarms for all the three small watersheds.

The KGE for 1-day lead daily rainfall forecast is below zero for all three small watersheds, indicating very poor performance in capturing daily watershed-averaged rainfall. In contrast, the KGE for the largest watershed was relatively high (KGE of 0.45), but still on the borderline between "poor" and "intermediate" skill.

The KGE increases with aggregating the forecasts over larger timescales. The increase in KGE performance is larger for Inga I (KGE increased from -0.01 at daily timescale to 0.56 at 15-day total timescale), while the increase for the remaining two small watersheds is still unsatisfactory (KGE increased from -0.32 to -0.03 for Ruzizi I, from -0.51 to -0.09 for Koni). The largest increase for Inga watershed was due to the large random error in the daily GFS forecast, which averages out at larger accumulation timescales, whereas the errors in the other two watersheds are mostly dominated by bias.

Therefore, we conclude that the performance of GFS in the Congo basin is largely a function of watershed area, location of the watershed, and accumulation timescales. GFS shows better performance for very large watersheds. In the case of small-sized watersheds, GFS shows low performance at daily timescales, however, its performance at larger accumulation timescales depends on the location of the watershed: for the Inga I watershed, located in the western part of the watershed, the KGE is 0.56 at 15-day rainfall aggregation timescale, but for the Ruzizi I and Koni watersheds located in the eastern and southern parts of the watershed, the KGE at 15-day timescale is still very low.

The performance metrics indicate that IMERG Early (satellite-inly rainfall product available in near-real time_ has much better performance compared to GFS, indicating that IMERG Early has the potential to be used in post-processing of GFS in order to improve the accuracy of GFS forecasts. We recommend exploring the possibility of using IMERG Early in post-processing of GFS forecasts, and for this purpose, a number of post-processing techniques could be explored, such as, simple bias (multiplicative) correction (Gumindoga et al. 2019), multi-resolution bias correction through wavelet analysis (Xu et al. 2019) or empirical mode decomposition method (Wang et al. 2020, Prasad et al. 2019), and Artificial Intelligence-based methods such as Feed Forward Neural Network (Cloud et al. 2019), Support Vector Machine ((Du et al. 2017; Yu et al. 2017), and Adaptive Neural Fuzzy Inference System (Jehanzaib et al. 2021).

CHAPTER 9

Evaluation of Medium-Range Forecasts (GFS) over Zambezi

9.1 Study Region

The Zambezi River Basin (Figure. 9.1), the fourth-longest river in Africa, has a basin area of 1.37 million km^2, and is shared by eight riparian countries: Zambia (42% of the basin area), Angola (18%), Zimbabwe (16%), Mozambique (12%), Malawi (7.5%), Tanzania (2%), Botswana (1.5%), and Namibia (1%) (Kling et al. 2014). Two large hydropower dams are located in the main Zambezi River, which are the Cohora Bassa Dam (2,075 MW) in Mozambique and further upstream in the Kariba Sam (2130 MW) at the border of Zambia and Zimbabwe. There is another dam, Itezhi-Tezhi Dam on the Kafue River, which is the most regulated tributary of the Zambezi River, that generates 120 MW. Table 9.1 provides more information about these dams. The elevation of the Zambezi basin ranges from 152 m above sea level to sea level. The basin lies in the tropics between 10 and 20°S, and is characterized by one distinct wet and one dry season. The movement of the Inter-Tropical Convergence Zone (ITCZ) in the Southern Hemisphere results in the peak rainy season that occurs during the summer (October to April) and the dry winter months (May to September) (Schlosser and Strzepek 2015).

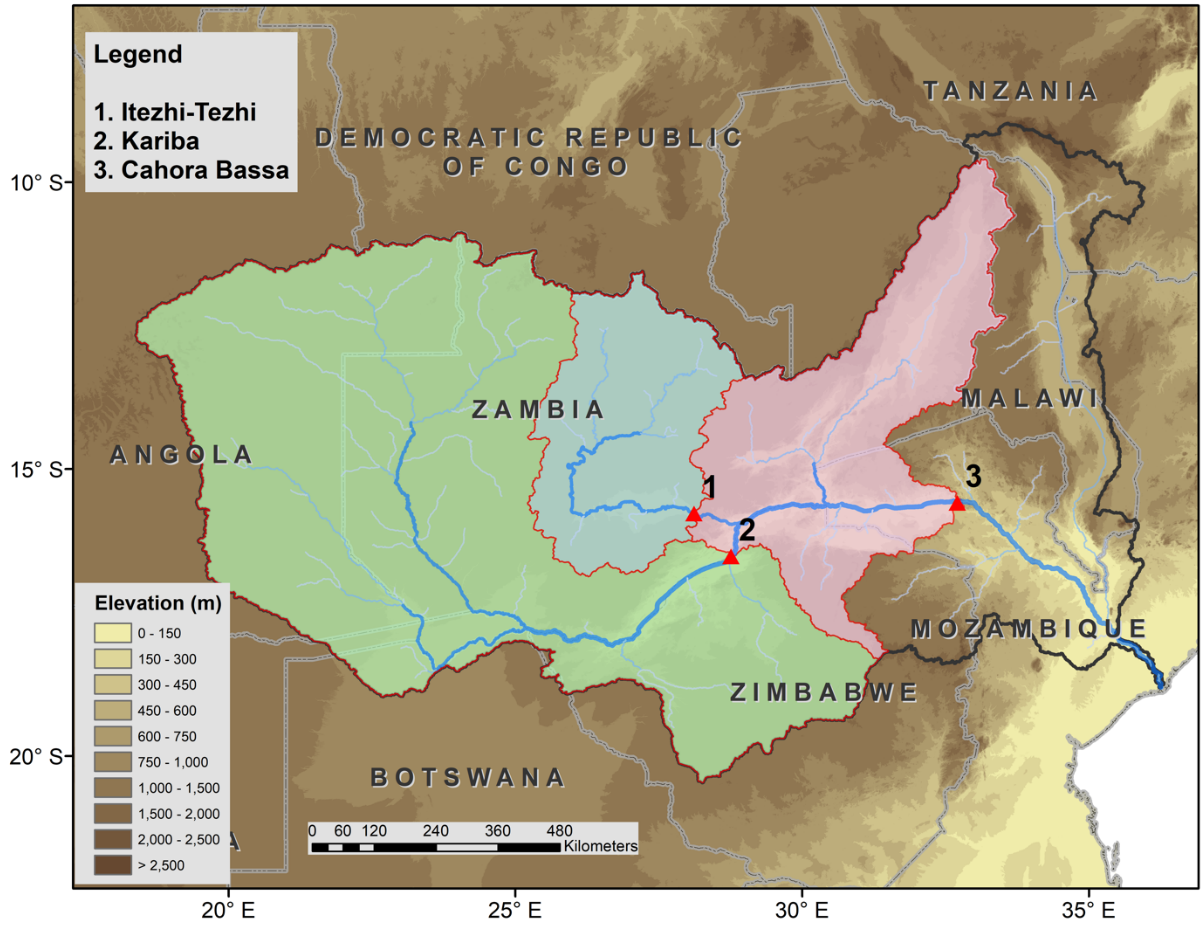


Figure 9.1. Map of Zambezi River Basin, with the location of selected dams/reservoirs: (1) Itezhi-Tezhi, (2) Kariba, and (3) Cahora Bassa, and the drainage basins defined by the dam locations (see the colored areas between dams).

Table 9.1. Major reservoir dams in the Zambezi River Basin

Dams	Country	Operational Since*	Capacity (million m^3)*	Power (MW)*	Purpose*			Area of Drainage Basins (km^2)**	Elevation of Drainage Basin (m)**
					Irrigation	Flood Control	Hydroelectricity		
Itezhi-Tezhi	Zambia	1978	4925	120			x	161815	1174
Kariba	Zambia	1959	94000	1626			x	729901	1113
Cahora Bassa	Mozambique	1974	52000	2075	x	x	x	250862	896

* information obtained from the Global Reservoir and Dam Database (Lehner et al. 2011) and Food and Agriculture Organization of the United Nations (FAO)'s Global Information System on Water and Agriculture (AQUASTAT).
** Calculated from HydroSEHDS (Lehner et al. 2008).

9.2 Results and Discussion

9.2.1 Annual Spatial Variability of Rainfall

Figure 9.2 presents the spatial map of annual (15 June 2019 to 15 June 2020) rainfall derived from the forecast and satellite products. According to the reference IMERG Final, most of the basin receives annual rainfall under 1,000 mm except for pockets of areas in the northwest and northeast. The other research-quality product, CHIRPS, show similar spatial pattern of rainfall as with IMERG Final. Validated against IMERG Final, the 1-day lead GFS shows large overestimation bias in much of the upper parts of the basin, but captures well rainfall in the lower part.

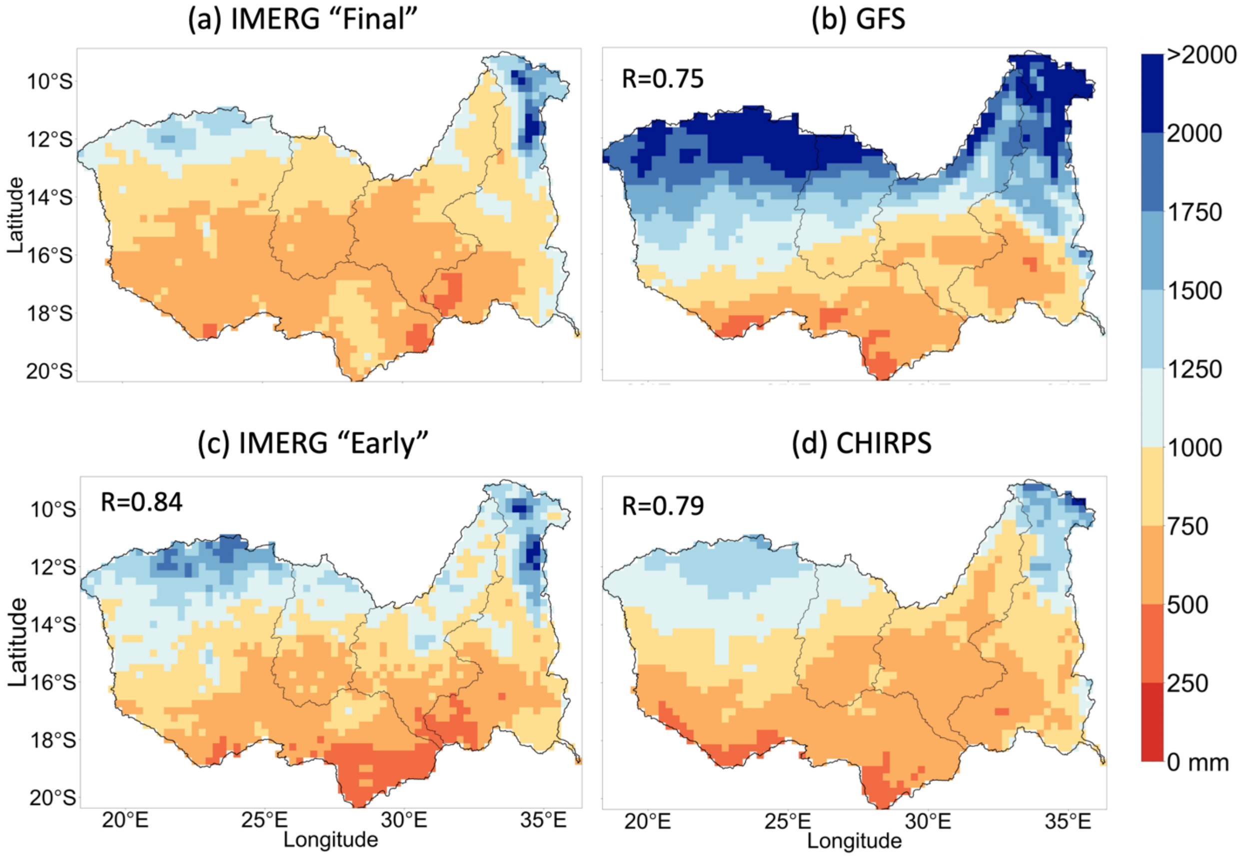


Figure 9.2. Spatial map of annual rainfall (in mm), for the period 15 June 2019 to 15 June 2020, derived from IMERG Final, GFS (1-day lead time), IMERG Early, and CHIRPS, for the Limpopo River Basin.

9.2.2 Monthly Cycle of Rainfall

Figure 9.4 presents the time series of monthly rainfall derived from the 1-day lead GFS and satellite precipitation products. According to IMERG Final, the region is characterized by a distinct wet season, November to March, and a dry season, April to October. The 1-day lead GFS forecasts captures the overall pattern but tends to overestimate monthly rainfall.

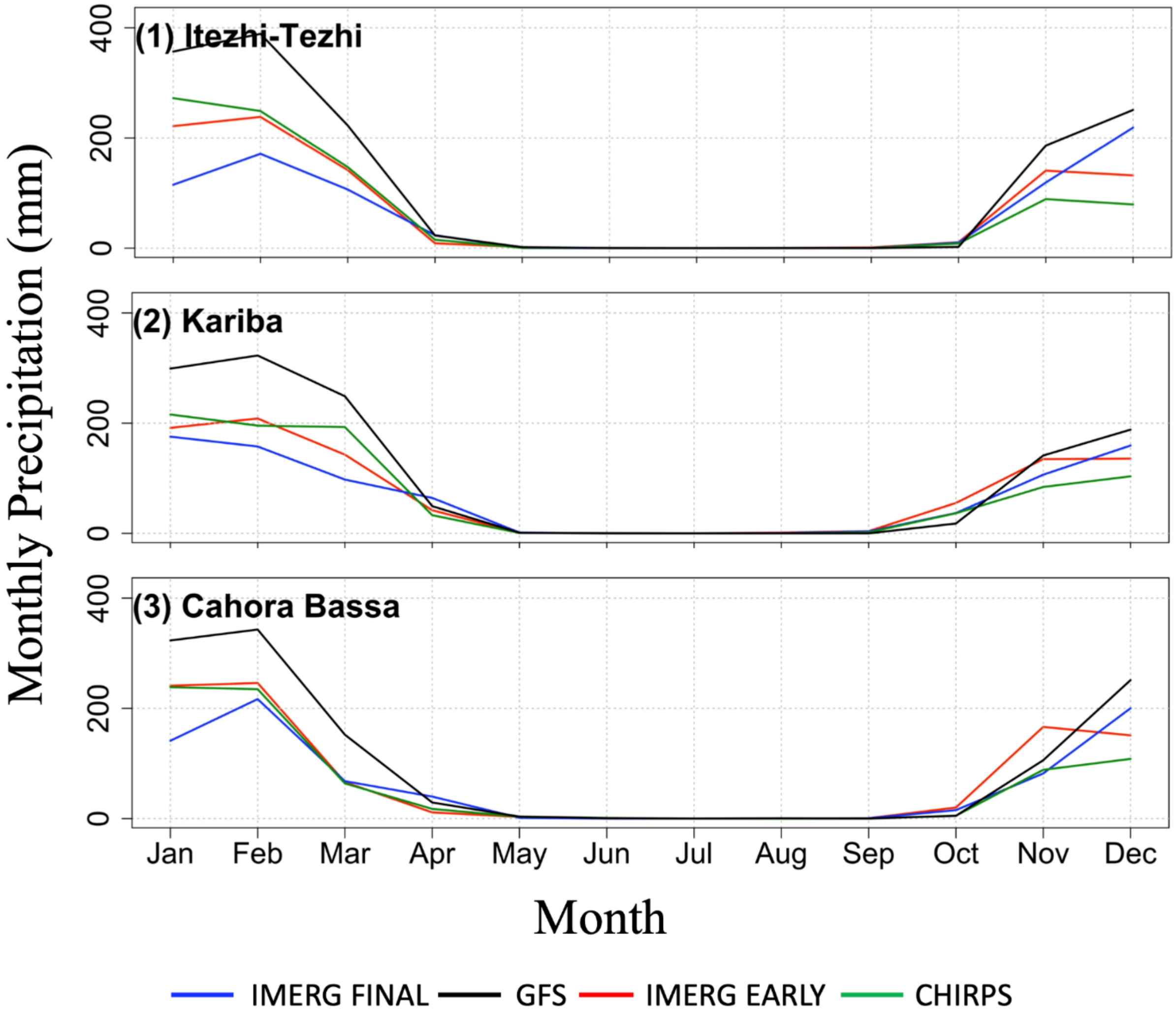


Figure 9.3 Monthly time series of sub-basin averaged precipitation (mm), for the period 15 June 2019 to 15 June 2020, derived from IMERG Final, GFS (1-day lead time), IMERG Early, and CHIRPS, for watersheds of three dams in the Zambezi River Basin: (1) Itezhi-Tezhi, (2) Kariba, and (3) Cahora Bassa.

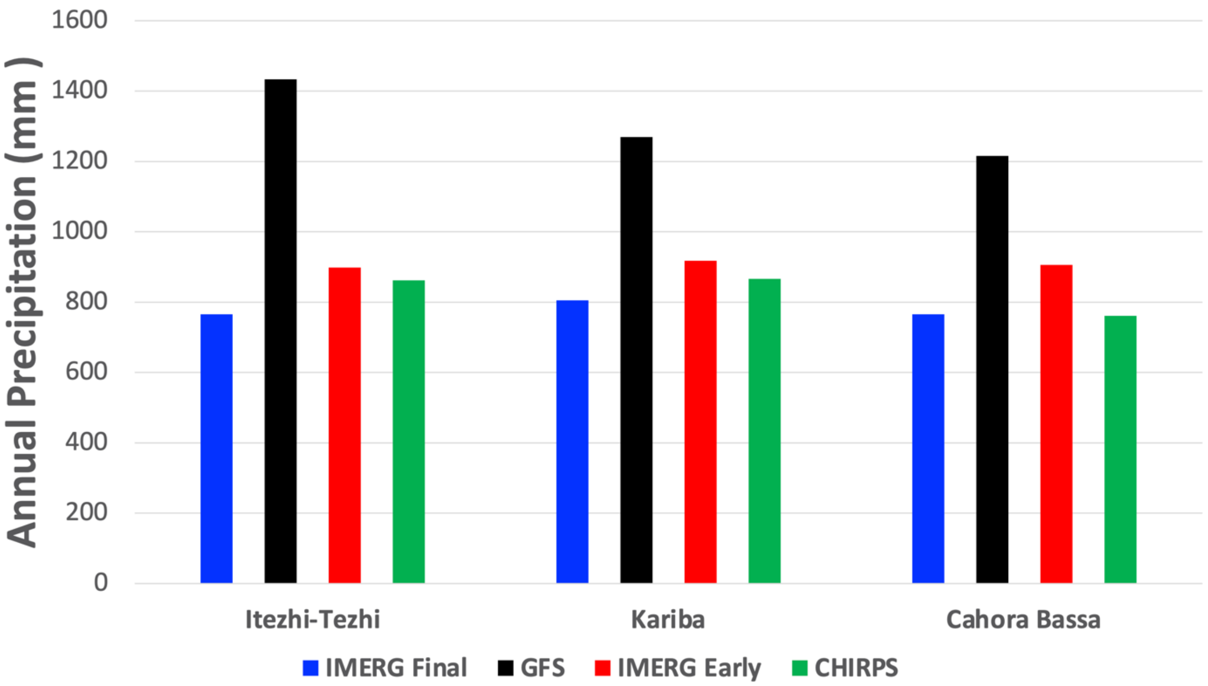


Figure 9.4. Sub-basin averaged annual precipitation (mm) for the period, 15 June 2019 to 15 June 2020, for each of the Congo's sub-basins, derived from the 1-day lead GFS forecast and different satellite precipitation products.

9.2.3 Annual Rainfall

Here, we aggregate the 1-day lead GFS forecasts and satellite rainfall products to annual time scale (Figure. 9.4). According to IMERG Final, the annual watershed-average rainfall is in the range 750 to 800 mm yr^{-1} in all three watersheds. The CHIRPS rainfall values are a bit higher and indicate annual rainfall in the range 760 to 860 mm yr^{-1} in all three watersheds. The 1-day lead GFS has large overestimation bias at all sites, with overestimation by 87% (66%), 58% (47%), 59% (60%), at the Itezhi-Tezhi, Kariba, and Cahora Bassa watersheds, respectively, compared to IMERG Final (CHIRPS) estimates.

9.2.4 Daily Time series

Figure 9.5 presents the daily time series of watershed-averaged rainfall, derived from 1-day lead GFS and satellite products. The GFS forecasts capture well the temporal variability of daily rainfall, but tend to overestimate the majority of daily rainfall at all sites.

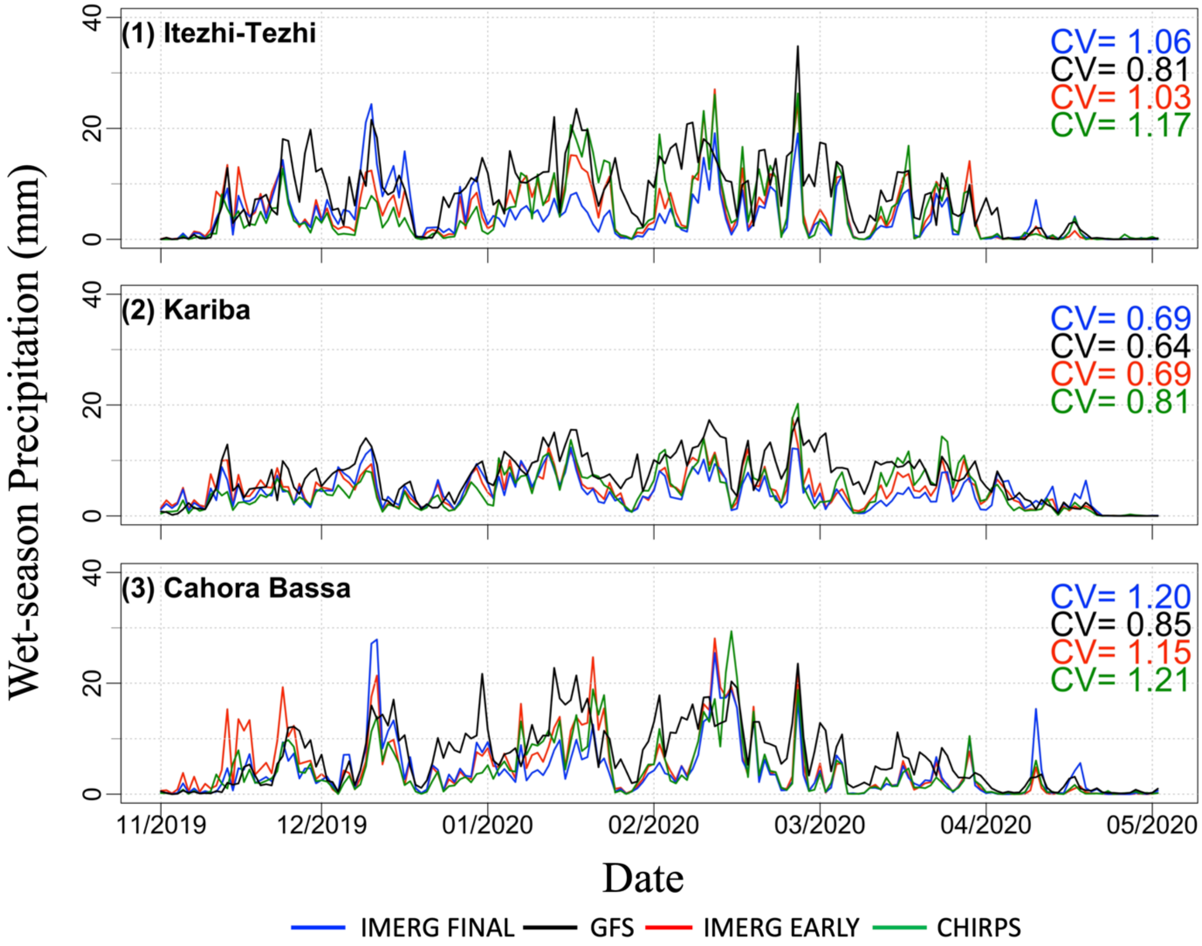


Figure 9.5. Time series of sub-basin averaged precipitation total (mm) for the wet period (November – April), for each of the dam watersheds, as derived from various precipitation products. The Figure also shows the coefficient of variation (CV) as a measure of temporal variation.

9.2.5 KGE Statistics

Figure 9.6 presents the performance statistics of various precipitation products (with respect to the IMERG Final estimates), in terms of Kling-Gupta Efficiency (KGE), Bias Ratio (BR), correlation (R), variability ratio (γ), and root mean square error normalized by reference precipitation mean (NRMSE). The KGE of 1-day lead GFS at all sites is within the range 0.00 to 0.27, which indicates “poor” skills in forecasting daily rainfall. The KGE can be decomposed into three statistics: bias ratio (BR), correlation (R), and variability ratio (γ). The GFS tends to overestimate daily rainfall by about 60% to 64% across the sites. The GFS has low correlation (R between 0.60 and 0.66) with IMERG Final, indicating that between 35% and 45% of the daily fluctuation in IMERG Final can be captured by GFS. The variability ratio for Itezhi-Tezhi and Cahora Bassa is about 0.75, and is close to one for Kariba. Therefore, the low KGE are primarily due to high bias and low correlation coefficient in the 1-day lead GFS forecasts. The root-mea-squared-error of GFS is in the range 100% to 152% of the mean rainfall rate.

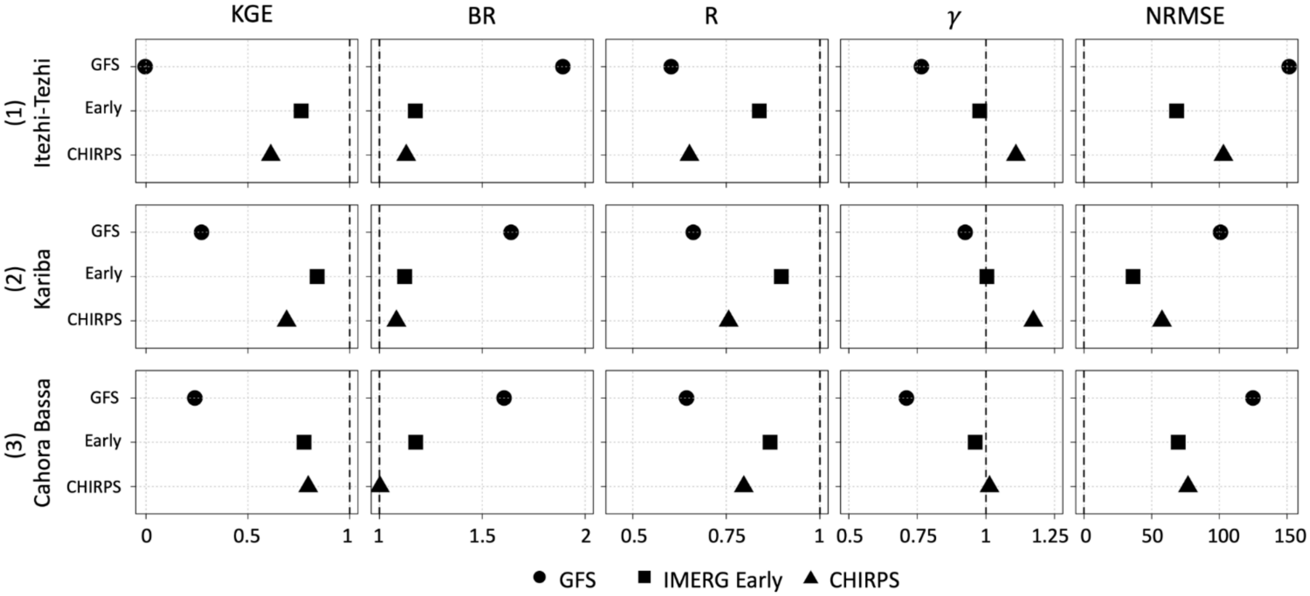


Figure 9.6 Summary of performance statistics (Kling-Gupta Efficiency KGE, Bias Ratio BR, correlation R, variability ratio γ, and root mean square error normalized by reference precipitation mean NRMSE) of 1-day lead GFS forecasts and different satellite products, for different dam watersheds in the Zambezi River Basin.

9.2.6 Dependence of Forecast Performance on Precipitation Rate

Figure 9.7 presents the scatterplot of 1-day lead GFS forecasts against daily IMERG Final rain rates. The 1-day lead GFS forecast tends to overestimate the majority of daily rainfall, at almost all rain rates.

9.2.7 Effect of Lead Time on Forecast Performance

Figure 9.8 presents the Kling-Gupta Efficiency (KGE) and its components for 5-day total rainfall forecast of GFS for three different lead time periods, 1-5 day, 6-10 day, and 10-15 day. As lead time increases, the KGE decreases. The KGE breakdown indicates that it is the correlation statistic (GFS with IMERG Final) that is mostly affected by the change in lead time. Similar findings are also obtained for a daily timescale (see Table 9.2).

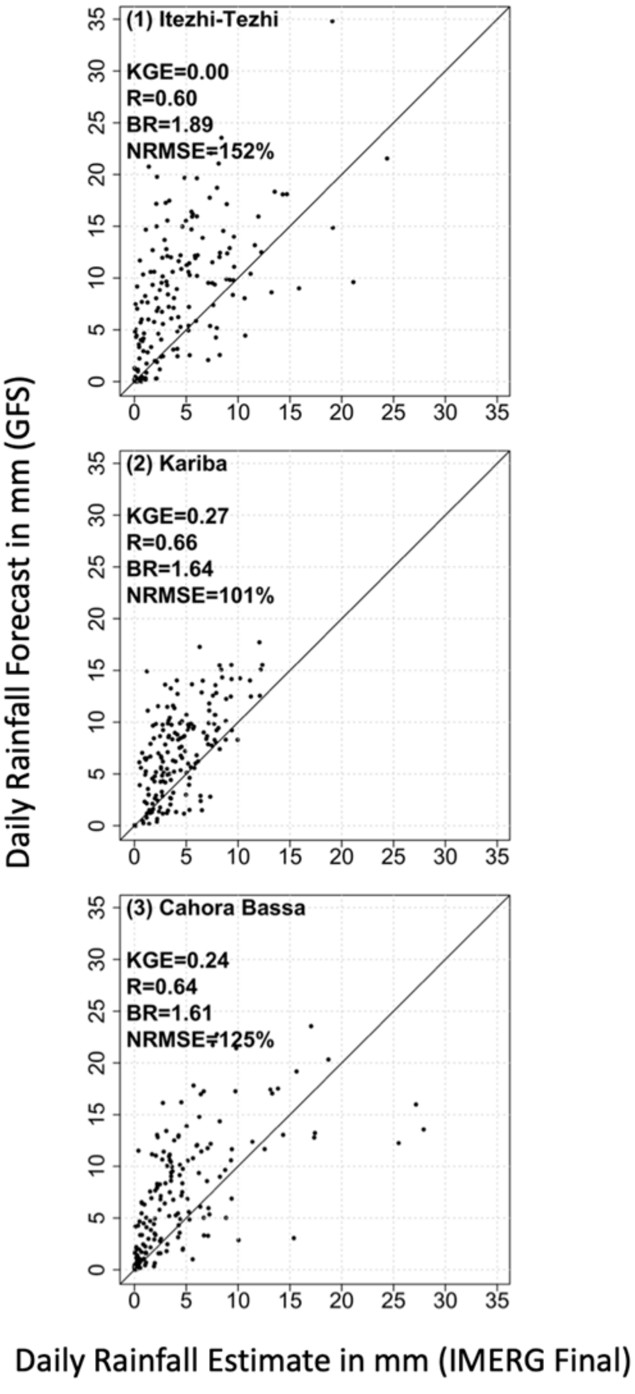


Figure 9.7 Scatterplot of watershed-averaged daily precipitation forecast obtained from 1-day lead GFS forecasts against corresponding values from IMERG Final, over each dam watershed in the Zambezi River Basin.

9.2.8 Effect of Temporal Aggregation Scale on Forecast Performance

Figure 9.9 presents the scatterplot of watershed-averaged 15-day accumulated rainfall obtained from the GFS forecasts against IMERG Final estimates, for each dam watershed. The 15-day total forecast is obtained by adding multiple lead-time daily forecasts: 1-day lead, 2-day lead, and all the way up to 15-day lead. Over all the watersheds, the 15-day accumulated rainfall shows slight increase in performance compared to the daily timescale (KGE changed from 0.0 at daily to 0.07 at 15-day accumulation for Itezhi-Tezhi, from 0.27 to 0.42 for Kariba, and from 0.24 to 0.23 for Cahora Bassa). We point out that aggregation over longer temporal scales involves the use of data from both short-lead and long-lead forecasts, a situation that leads to two contradictory error natures. On the one hand, introducing data from long-lead times brings larger errors (see Section 3.7) but, on the other hand, the temporal aggregation tends to average out the errors. This study shows that the effect of averaging out the errors outweighs the additional error introduced by the use of long-lead forecast.

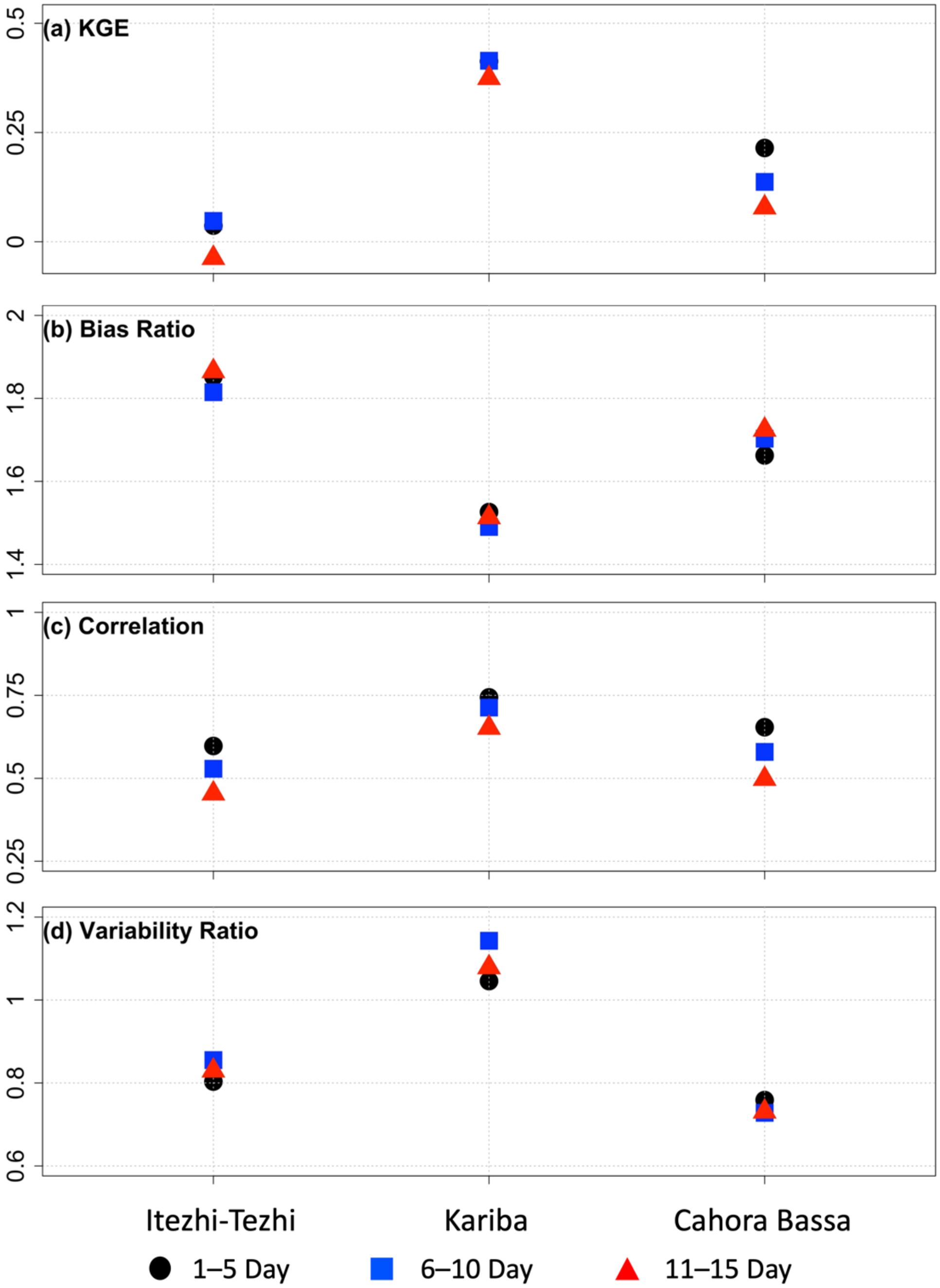


Figure 9.8. Kling-Gupta Efficiency (KGE) and its components for 5-day total rainfall forecast of GFS for three different lead time periods, 1-5 day, 5-10 day, and 10-15 day, for each dam watershed.

9.2.9 Comparison of the Performance of IMERG Early and GFS

The poor skills of 1-day lead GFS at all three watersheds suggest the need for exploring ways of improving the GFS forecast accuracy. One such method is post-processing of GFS forecasts using rainfall estimates that have relatively better accuracy and are available in near-real time. Satellite-only products, such as IMERG Early, are available in near-real time. In this section, we explore how the performance of IMERG Early compare with the performance of GFS. The spatial distribution of IMERG Early has better correlation with IMERG Final (R = 0.84), compared to GFS, and does not suffer from the large overestimation bias in the upper part of the watershed like GFS (Figure. 9.2). In terms of monthly and annual rainfall values (Figures. 9.3 and 9.4), IMERG Early outperforms GFS over all three watersheds as it has much lower bias compared to GFS. IMERG Early has KGE values around 0.75, which is much higher than the KGE of 1-day lead GFS (Figure. 9.6). The higher KGE for IMERG Early is the result of its better bias and correlation with respect to IMERG Final. Therefore, all performance measures indicate that the IMERG Early has better performance than the GFS, which implies that the IMERG Early products could be considered for post-processing of GFS in order to improve the accuracy of GFS forecasts.

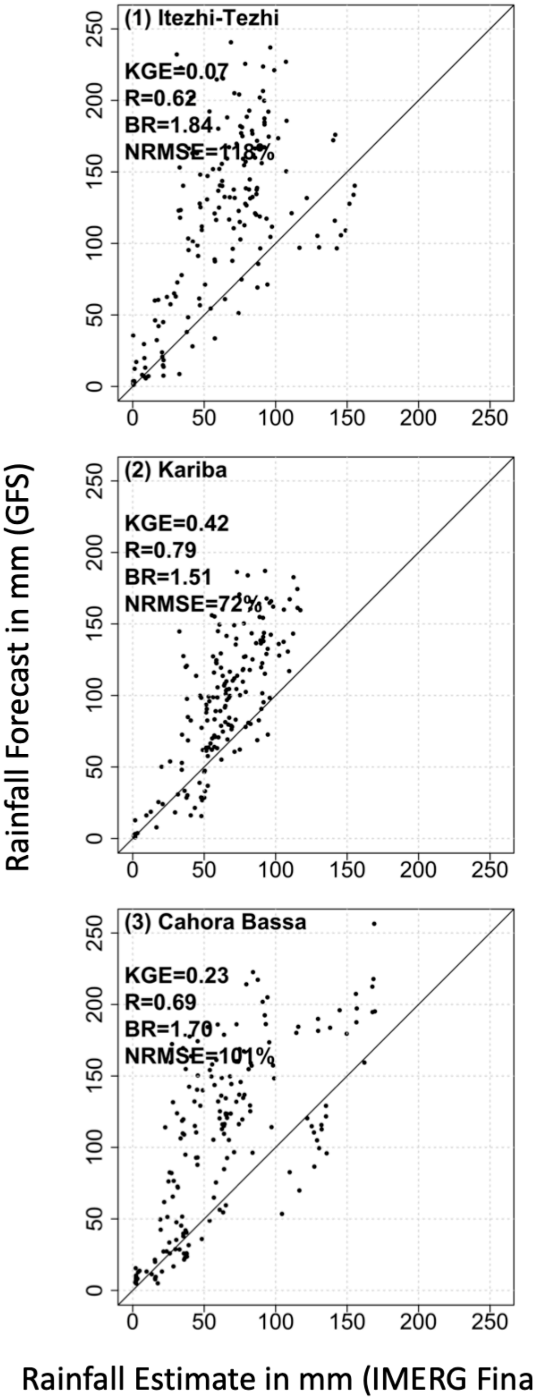


Figure 9.9 Scatterplot of watershed-averaged 15-day accumulated precipitation forecast obtained from GFS forecasts against corresponding values from IMERG Final, over each dam watershed in the Zambezi River Basin.

9.2.10 Performance of GFS if the Reference Product is Changed from IMERG Final to CHIRPS

We acknowledge that the reference dataset used in our evaluation (i.e., IMERG Final) has its own estimation errors. We conducted additional assessment to evaluate the performance of GFS using CHIRPS rainfall products as reference. Table 9.2 shows the performance statistics of GFS for different lead times, using IMERG FINAL and CHIRPS, separately, rainfall products as reference. The overall magnitude of GFS performance (as well as the variability of the performance across watersheds) is similar when either rainfall product is used as a reference. Both reference products reveal “poor” skills for 1-day lead GFS forecast for all three watersheds, with KGE ranging from 0.0 to 0.27 with IMERG Final as reference, and KGE from 0.19 to 0.38, with CHIRPS as reference. The GFS performance with IMERG Final as reference is found to be slightly worse than with CHIRPS as reference. Therefore, our results show that the overall performance of GFS remains about the same if the reference product were to be changed from IMERG Final to CHIRPS, indicating the robustness of IMERG Final as reference product.

Table 9.2 Performance statistics of daily GFS forecast for various lead times (1-day, 5-day, 10-day, and 15-day) using IMERG Final (CHIRPS) rainfall products as reference, in terms of correlation, bias ratio, and NRMSE.

Lead time of GFS forecast	Correlation	Bias Ratio	KGE	NRMSE (%)
		Itezhi-Tezhi		
1-day	0.60 (0.68)	1.89 (1.67)	0.00 (0.19)	152 (123)
5-day	0.39 (0.48)	1.78 (1.57)	-0.03 (0.14)	154 (132)
10-day	0.42 (0.50)	1.81 (1.60)	-0.04 (0.14)	156 (134)
15-day	0.28 (0.31)	1.93 (1.71)	-0.22 (-0.06)	178 (160)
		Kariba		
1-day	0.66 (0.73)	1.64 (1.52)	0.27 (0.38)	101 (84)
5-day	0.58 (0.64)	1.44 (1.33)	0.39 (0.47)	88 (77)
10-day	0.41 (0.53)	1.47 (1.36)	0.25 (0.39)	107 (92)
15-day	0.31 (0.28)	1.55 (1.43)	0.12 (0.13)	115 (112)
		Cahora Bassa		
1-day	0.64 (0.71)	1.61 (1.60)	0.24 (0.27)	125 (116)
5-day	0.48 (0.56)	1.68 (1.68)	0.08 (0.12)	146 (137)
10-day	0.45 (0.51)	1.76 (1.76)	-0.01 (0.02)	154 (147)
15-day	0.31 (0.38)	1.73 (1.73)	-0.07 (-0.03)	165 (159)

9.3 Conclusions

In this study, medium-range (1-day to 15-day lead time) precipitation forecasts available from the Global Forecast System (GFS) have been evaluated in the Zambezi River Basin, focusing on some of its major dams: Itezhi-Tezhi, Kariba, and Cahora Bassa. The evaluation is done using the satellite-gauge merged rainfall product IMERG Final as reference, and additional analysis is conducted by changing the reference product to CHIRPS. Additionally, the performance of GFS is compared to the performance of the near-real time, satellite-only rainfall product, IMERG Early, to see if IMERG Early has the potential to be used in post-processing of GFS.

Our results indicate that the 1-day lead GFS overestimates rainfall over the vast majority of the basin. For the three dam watersheds considered, the 1-day lead GFS overestimates rainfall by 89%, 64%, and 61%, for Itezhi-Tezhi, Kariba, and Cahora Bassa, respectively. The correlation coefficient between the 1-day lead, daily, GFS forecast and IMERG Final is in the range 0.60 to 0.66, for all three watersheds. The skill of the 1-day lead GFS forecast is poor (KGE = 0 to 0.27) at all three dam watersheds, indicating the need to improve the accuracy of these forecasts. The breakdown of KGE indicates that the low KGE value is due to the high bias of GFS. Aggregating the forecasts from daily to longer timescale slightly increases the GFS forecast accuracy. IMERG Early has better performance than GFS over all three watersheds, primarily because IMERG Early has low bias compared to GFS. Therefore, we conclude that the accuracy of GFS forecasts in the Zambezi River Basin is low, mostly due to high bias. Given that IMERG Early outperforms GFS, we recommend testing the suitability of IMERG Early to serve as input into post-processing of GFS in order to improve the accuracy of GFS forecasts.

CHAPTER 10

Evaluation of Medium-Range Forecasts (GFS) over Limpopo

10.1 Study Area

The Limpopo River Basin (Figure 10.1) has a drainage area of 415,000 km^2 and a river length of 1,750 km, and is shared by four countries: South Africa (45% by area), Botswana (20%), Mozambique (20%), and Zimbabwe (15%). The river originates from the highlands that separate South Africa from Botswana and Zimbabwe and flows through and between these countries before it enters Mozambique and finally drains into the Indian Ocean (FAO 2004). It serves as a border for about 640 km, separating South Africa to the southeast from Botswana to the northwest and Zimbabwe to the north. The terrain of the basin consists largely of an undulating landscape of plains, punctuated with ranges of hills and mountains (CGIAR 2014).

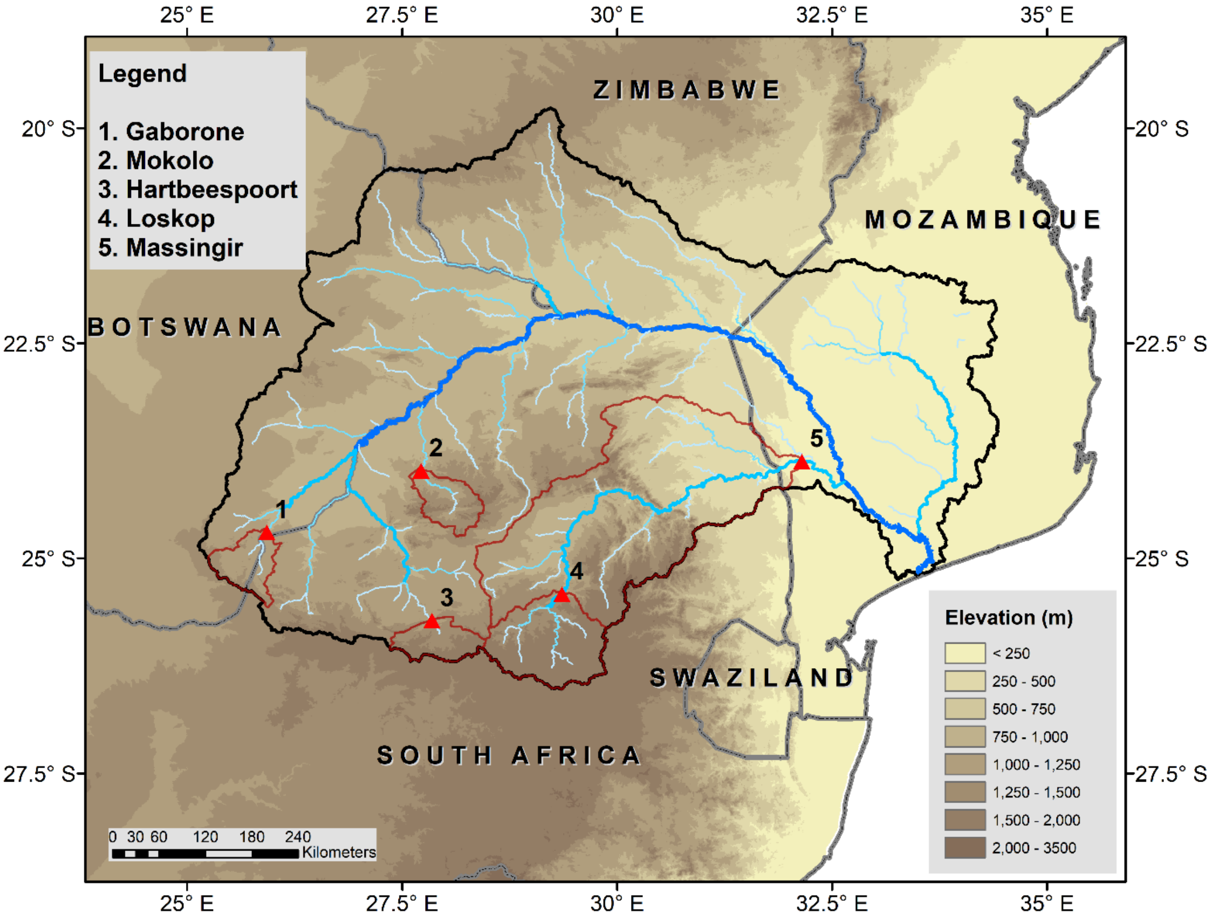


Figure 10.1. Map of Limpopo River Basin, with the location of selected dams/reservoirs: (1) Gaberone, (2) Mokolo, (3) Hartbeespoort, (4) Loskop, and (5) Massingir, and the drainage basins defined by the dam locations (see the red watershed boundary).

There are many dams in the Limpopo River Basin. For this study, we selected five dams located in different parts of the basin (Figure 10.1, Table 10.1). The selected dams have drainage areas ranging from 4,991 km^2 (Hartbeespoort Dam) to 68,779 km2 (Massingir). The Massingir Dam in Mozambique has the largest storage capacity in the basin at 2,800 Mm^3, and is a multi-purpose irrigation and hydropower project (Barros 2009). The Loskop Dam in South Africa is the second largest dam in the basin, with a storage capacity of 375 Mm^3. The Mokolo Dam (also known as the Hans Strijdom Dam) as has a storage capacity of has a capacity of 145 Mm^3, and mainly serves for municipal and industrial purposes. The Gaborone dam in Botswana has a storage capacity of 144 Mm³, and supplies water to the capital city of Gaborone.

Table 10.1. Major reservoir dams in the Limpopo River Basin

Dams	Country	Operational Since*	Capacity (million m^3)*	Power (MW)*	Purpose*			Area of Drainage Basins (km^2)**	Elevation of Drainage Basins (m)**
					Irrigation	Flood Control	Hydroelectricity		
Gaborone	Botswana	1985	144		x			5,353	1,183
Mokolo	South Africa	1980	146		x			5,195	1,305
Hartbeespoort	South Africa	1969	194.6		x			4,991	1,492
Loskop	South Africa	1979	361		x			15,162	1,544
Massingir	Mozambique	1976	2,256	60	x	x	x	68,779	910

* information obtained from the Global Reservoir and Dam Database (Lehner et al. 2011) and Food and Agriculture Organization of the United Nations (FAO)'s Global Information System on Water and Agriculture (AQUASTAT).
** Calculated from HydroSEHDS (Lehner et al. 2008).

10.2 Results and Discussion

10.2.1 Annual Spatial Variability and Seasonal Characteristics

Figure 10.2 presents the spatial map of annual (15 June 2019 to 15 June 2020) rainfall derived from the 1-day lead GFS forecast and satellite products. According to the reference rainfall (i.e. IMERG Final), the annual rainfall varies mostly between 300 mm and 800 mm, with an increasing rainfall gradient in the north-south direction, with 300 to 400 mm in the north to 700 to 800 mm in the south. The 1-day lead GFS forecast also shows a north-south gradient, but there are clear differences in magnitude from the reference rainfall observations. GFS tends to underestimate in the northern (dry) part of the watershed, but overestimate in the southernmost (wet and mountainous) parts of the watershed. The overall correlation between the spatial map of annual GFS forecast and IMERG Final is 0.70, indicating that GFS can only capture about half of the spatial pattern in IMERG Final. As far as the performance of the other satellite rainfall products is concerned, the near-real-time product IMERG Early captures the overall spatial pattern better than the GFS forecast especially in the western part of the watershed, but tends to have comparable bias with IMERG Early. The other research-quality, satellite-gauge products, CHIRPS show lower spatial correlation with IMERG Final (correlation of 0.54), mainly in the eastern part (relatively low-elevation) of the watershed.

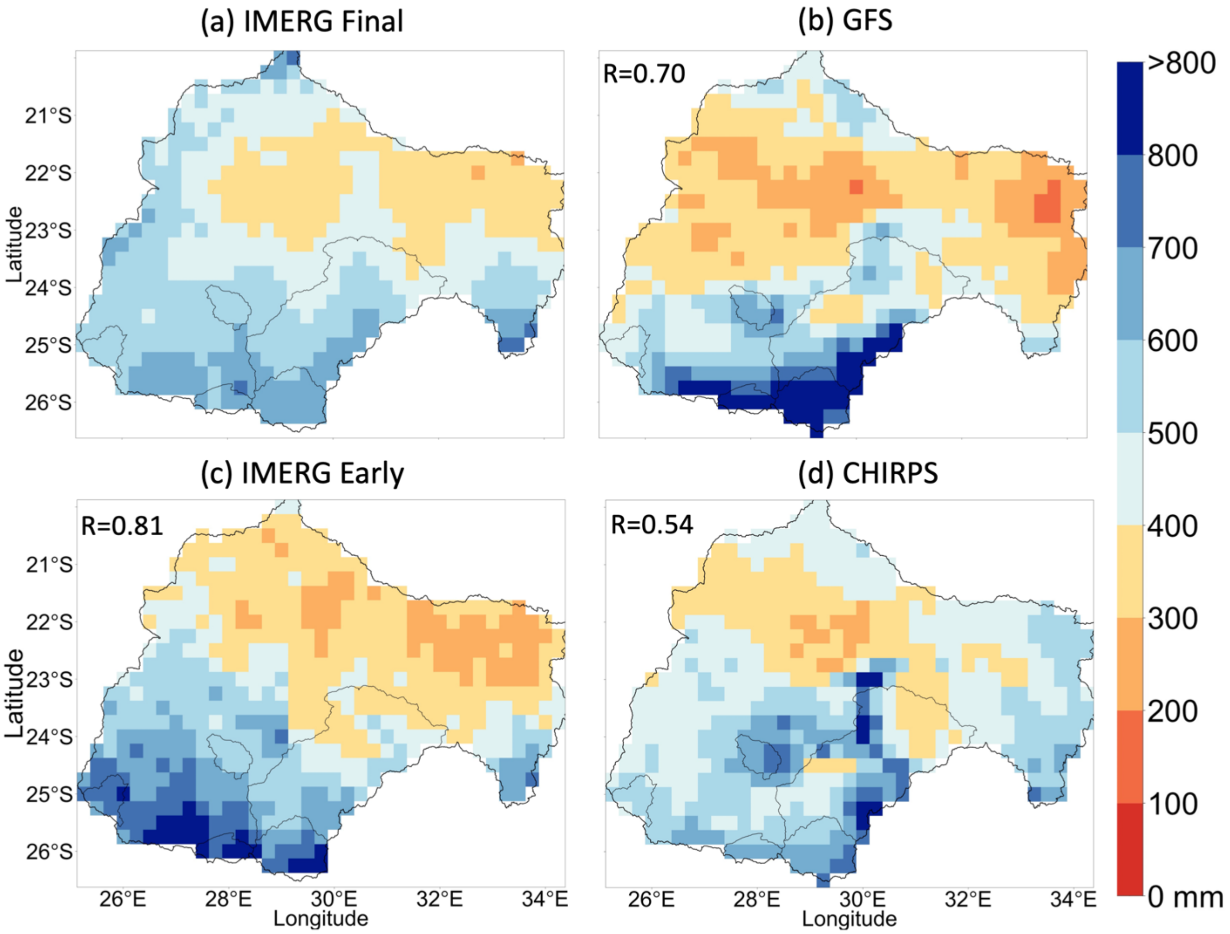


Figure 10.2. Spatial map of annual rainfall (in mm), for the period 15 June 2019 to 15 June 2020, derived from IMERG Final, GFS (1-day lead time), IMERG Early, and CHIRPS, for watersheds of five dams: (1) Gaborone, (2) Mokolo, (3) Hartbeespoort, (4) Loskop, and (5) Massingir.

Figure 10.3 presents the monthly time series of watershed-averaged rainfall derived from the 1-day lead GFS and three satellite precipitation products, for each of the dam watersheds. The rainfall season lasts six months, from November to April. In the three dam watersheds (Mokolo, Hartbeespoort, and Loskop), which are characterized by mountainous terrain and relatively large rainfall, GFS forecasts have large overestimation bias. In the two dam watersheds (Gaborone and Massingir), which are characterized by medium-elevation terrain and moderate rainfall, GFS has relatively low bias in almost all months. IMERG Early performs better than

GFS in capturing monthly rainfall in almost all watersheds and months, except in November, where IMERG Early has large overestimation bias in all cases. CHIRPS estimates are almost identical to IMERG Final, except for the Mokolo watershed, where CHIRPS gives much higher rainfall than IMERG Final.

10.2.2 How Well is the Time Series of Daily Precipitation Forecasted?

Figure 10.4 presents the time series of watershed-averaged daily rainfall, for the wet period November – April, for each of the dam watersheds. According to IMERG Final, the temporal variability (as measured through coefficient of variation or CV) of watershed-averaged daily rainfall varies from 1.54 to 2.41. The GFS tends to overestimate CV in Gaborone and Mokolo and underestimate CV in Hartbeespoort and Loskop, except for the largest watershed (Massingir) where GFS provides similar CV as IMERG Final. In terms of CV, IMERG Early outperforms GFS in almost all subbasins but is outperformed by GFS in Massingir. As far as CHIRPS is considered, its CV estimates are higher than IMERG Final in all subbasins.

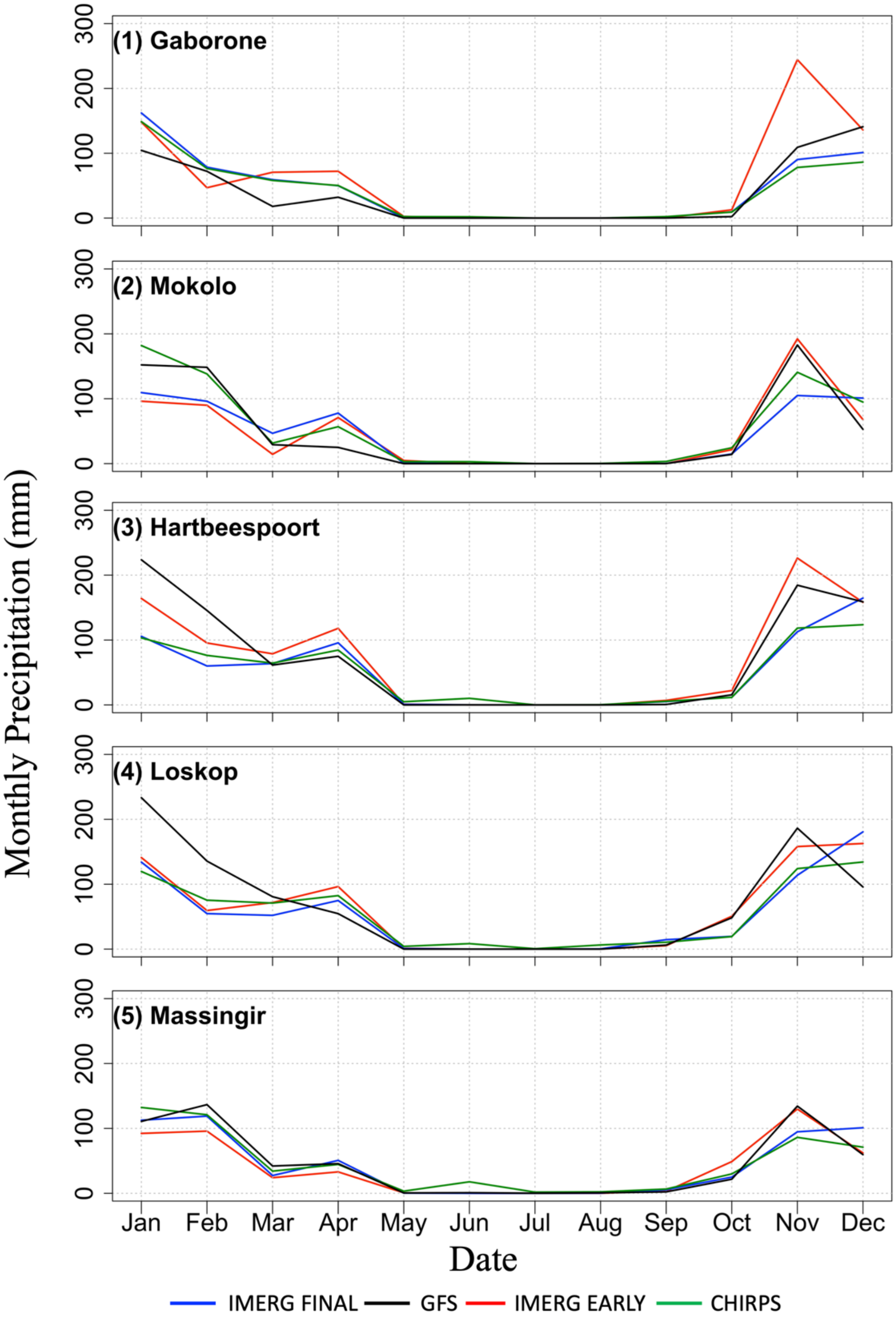


Figure 10.3. Monthly precipitation regime for the five watersheds of the Limpopo river Basin: (1) Gaborone, (2) Mokolo, (3) Hartbeespoort, (4) Loskop and (5) Massingir. Analyses are based on rainfall fields derived from IMERG Final, 1-day-lead GFS. IMERG Early, and CHIRPS. The time period covers from 15 June 2019 to 15 June 2020.

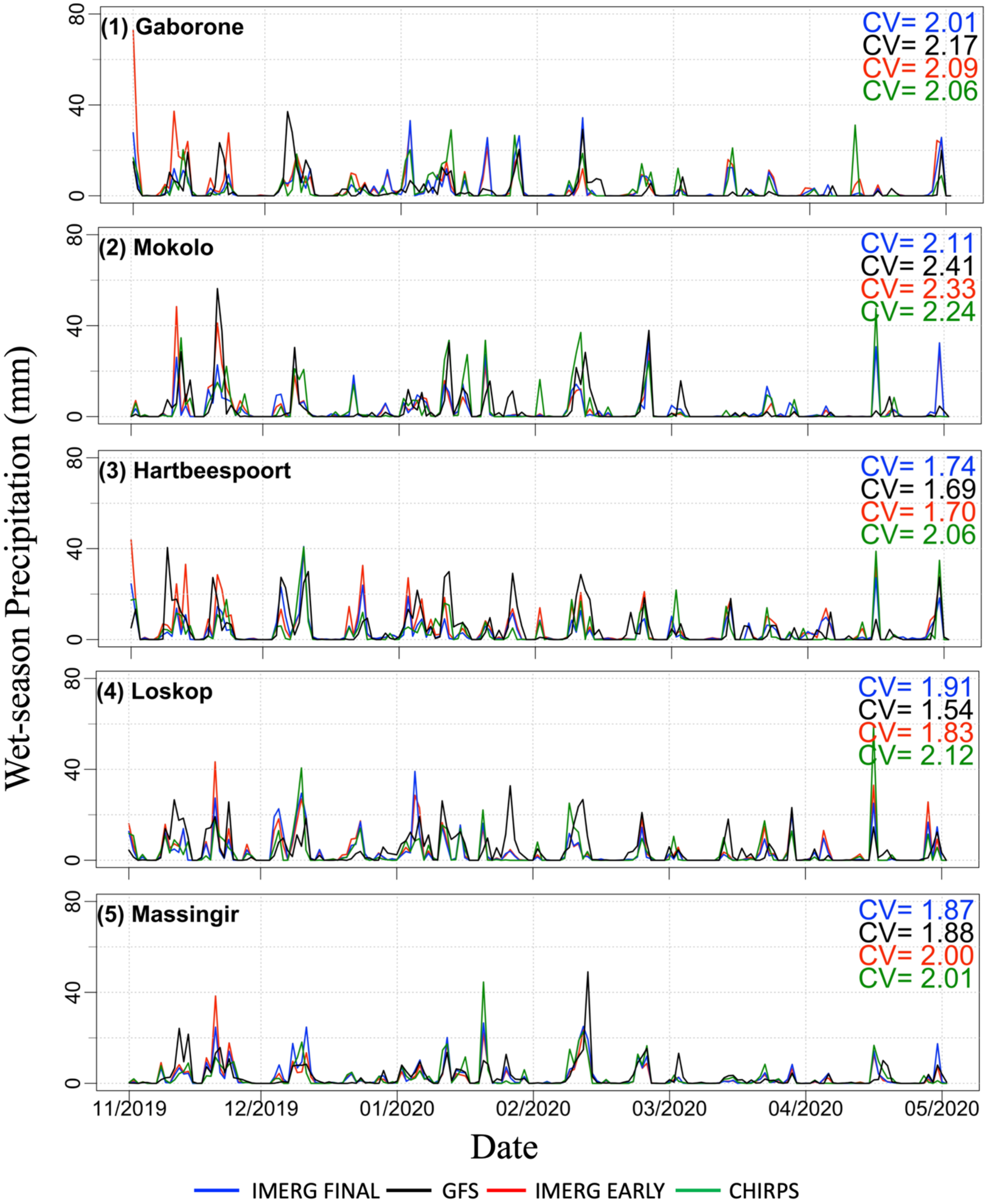


Figure 10.4. Time series of sub-basin averaged precipitation total (mm) for the wet period (November – April), for each of the dam watersheds, as derived from various precipitation products. The Figure also shows the coefficient of variation (CV) as a measure of temporal variation.

10.2.3 Dependence of Forecast Performance on Precipitation Rate

Figure 10.5 presents the scatterplot of watershed-averaged daily rainfall obtained from the 1-day lead GFS forecast against IMERG Final estimates, for each dam watershed. The 1-day lead GFS forecast has bias ranging from -12% (Gaborone), to 4% (Massingir), to 9% (Mokolo), to 29% (Loskop), and to 40% (Hartbeespoort). The bias is worse for the two highly mountainous watersheds, Loskop and Hartbeespoort. In all watersheds, the bias is mostly positive for light and moderate rainfall, but negative for heavy rainfall. The correlation between the GFS forecast and IMERG Final is about 0.50 for all watersheds, except for the Massingir watershed where the correlation jumps to 0.66. The highest correlation for the Massingir watershed could be partly attributed to its large watershed size. The KGE is about 0.50 (borderline between “poor” and “intermediate” skill) for all watersheds, except for the Massingir watersheds where it jumps to 0.66 (“intermediate” skill).

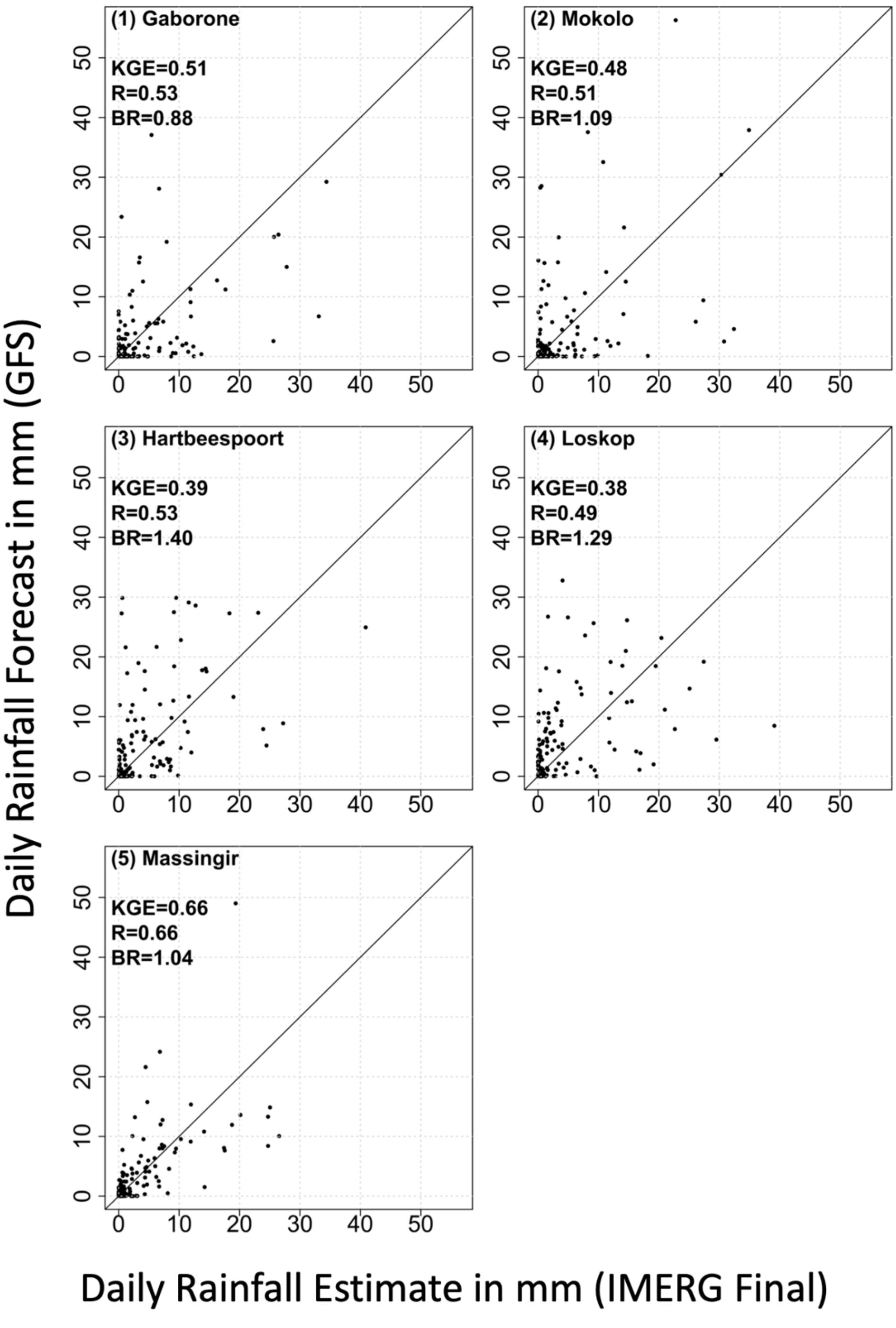


Figure 10.5. Scatterplot of watershed-averaged daily precipitation forecast obtained from the 1-day lead GFS forecast against corresponding values from IMERG Final, for five dam watersheds.

10.2.4 What is the Effect of Accumulation Timescale on the Forecast Performance?

Figure 10.6 presents the scatterplot of watershed-averaged 15-day total rainfall obtained from the GFS forecasts against IMERG Final estimates, for each dam watershed. The 15-day total forecast is obtained by adding multiple lead-time daily forecasts: 1-day lead, 2-day lead, and all the way up to 15-day lead. For the two highly mountainous watersheds, the 15-day total GFS forecast has the following performance statistics: high overestimation bias (66% at Hartbeespoort and 70% at Loskop), almost insignificant correlation coefficient (0.07 at Hartbeespoort and 0.25 at Loskop), and very poor skill (KGE < 0). The performance statistics for the medium-elevation watersheds are a bit higher: bias of 13% and 33%, correlation of 0.24 and 0.46, and KGE of 0.05 and 0.19, at Gaborone and Mokolo watersheds, respectively. For the largest watershed, Massingir, the performance statistics improve to a bias of 23%, correlation of 0.56, and KGE of 0.5 (borderline between “poor” and “intermediate” skill). In all cases, the 15-day total GFS forecast shows inferior performance compared to the daily GFS forecast. We point out that aggregation over longer timescales (from daily to 15-day total) is subject to two different error natures: effect of averaging on forecast errors, and effect of lead time on forecast errors. We anticipate that the effects of lead time on errors to dominate as averaging over longer time scale is expected to reduce random errors.

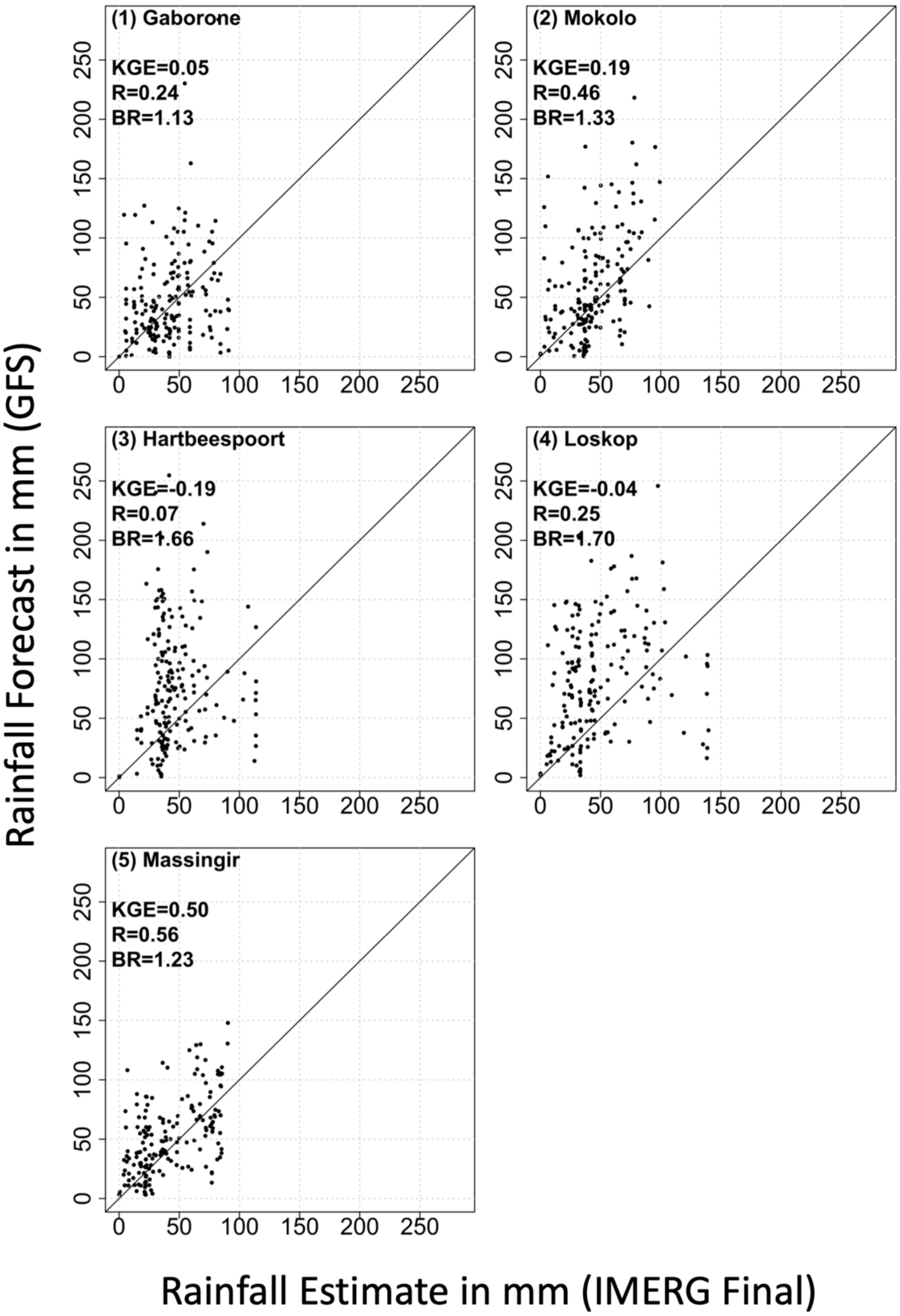


Figure 10.6. Same as Figure 10.5 but for 15-day total rainfall.

10.2.5 What is the Effect of Lead Time on the GFS Forecast Performance?

In Figure 10.7, we show the Kling-Gupta Efficiency (KGE) and its components for 5-day total rainfall forecast of GFS for three different lead time periods, 1-5 day, 6-10 day, and 10-15 day. The KGE values are much higher at short lead times (1 to 5 day) than at longer lead times (6 to 10 day, and 11 to 15 day). The KGE values at short lead times are higher by about 0.50 compared to the longer lead times. The KGE values fall sharply at 5-day lead time, and KGE does not change much beyond a lead time of 5 days. The breakdown of the KGE scores (BR, R and γ) reveals the key factors contributing to the differences in KGE estimates across lead times. The main differences in KGE statistics across the lead times can be primarily attributed to the correlation statistics: the correlation between the 6-10 day (or the 11-15 day) and the IMERG Final is under 0.25 in all watersheds, but jumps to 0.50 to 0.75 for the correlation between 1-5 day and IMERG Final, in all watersheds. Therefore, in order to improve the accuracy of GFS forecasts at longer lead times, the temporal fluctuation of GFS forecasts needs to be improved further to match that for IMERG Final.

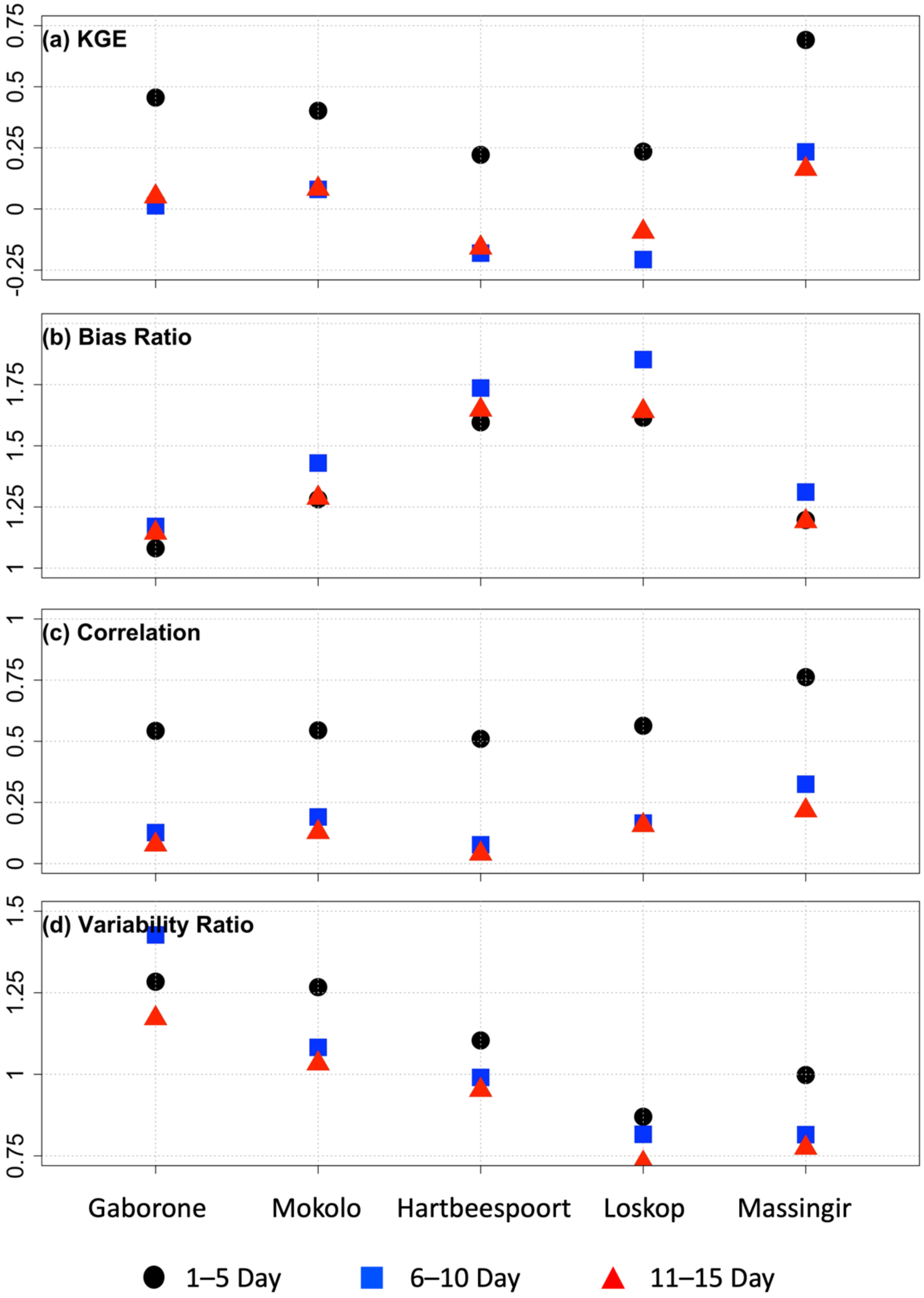


Figure 10.7. Kling-Gupta Efficiency (KGE) and its components for 5-day total rainfall forecast of GFS for three different lead time periods, 1-5 day, 5-10 day, and 10-15 day, for each dam watershed.

10.2.6 Comparison of the Performance of IMERG Early and GFS

The low performance statistics of GFS indicate the need for calibrating GFS forecasts in order to improve their accuracy. One way to achieve this could be through the use of post-processing techniques that involve rainfall estimates that have relatively better accuracy and are available in near-real time. Here, we assess the performance of IMERG Early, and compare it with the performance of GFS. Figure 10.8 presents the performance statistics of daily and watershed-averaged derived from IMERG Early, CHIRPS, and 1-day lead GFS forecast (with respect to the IMERG Final estimates), in terms of Kling-Gupta Efficiency (KGE), Bias Ratio (BR), correlation (R), variability ratio (γ), and root mean square error normalized by reference precipitation mean (NRMSE). The KGE values for IMERG Early (0.57 to 0.86) are higher than those for GFS (0.66 to 0.38) in all watersheds. The correlation statistics for IMERG Early are very high (between 0.75 and 1) compared to GFS (0.50 to 0.75), and this is expected due to the similarity of the algorithm of the two IMERG products. In terms of bias ratio, IMERG Early outperforms GFS in some watersheds (Mokolo and Loskop), but is outperformed by GFS in other watersheds (Gaborone) or has similar performance as GFS (Massingir). In terms of the variability ratio, IMERG Early outperforms GFS in some watersheds (Mokolo, Loskop, and Gaborone), or is outperformed by GFS (Massingir), or has similar value as GFS (Hartbeespoort). As far as NRMSE is considered, the RMSE of daily GFS is about 200% of the IMERG mean rainfall rate for four watersheds and is about 150% for the Massingir watershed. In comparison, IMERG Early's RMSE is about 100% of the IMERG mean rainfall rate at four watersheds and is about 170% at the Gaborone watershed.

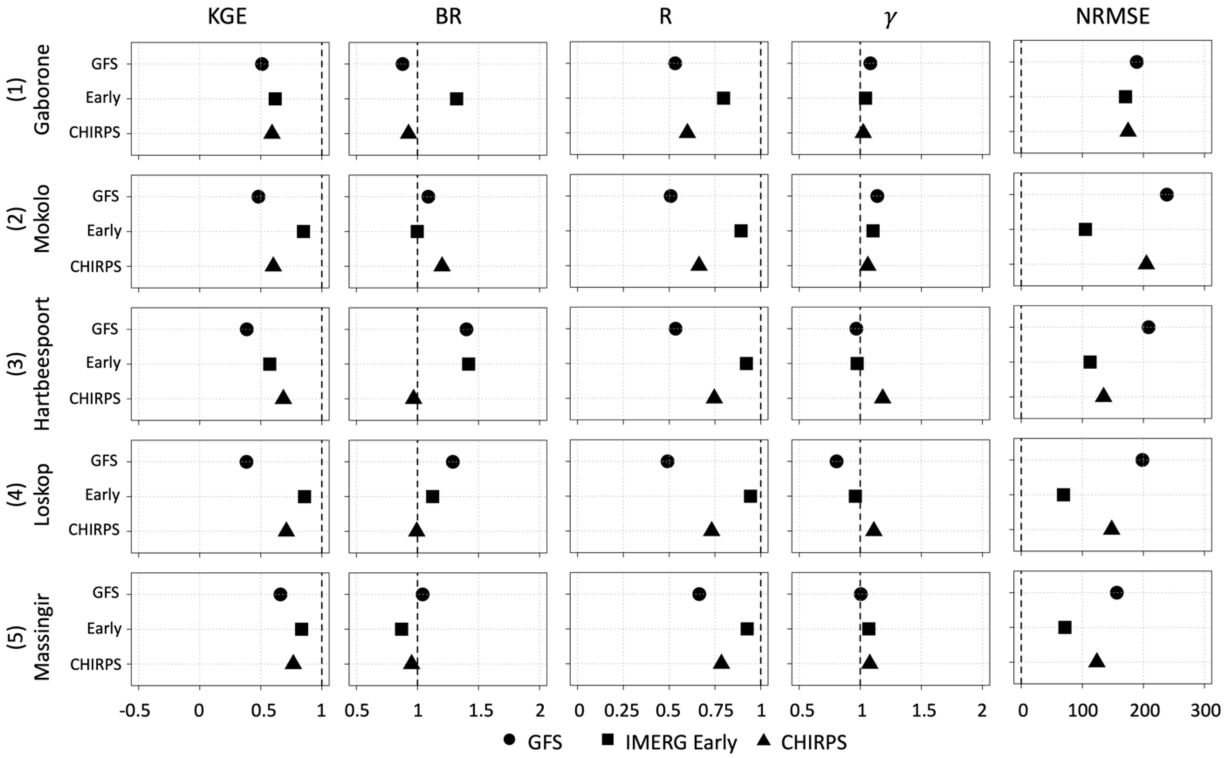


Figure 10.8. Summary of performance statistics (Kling-Gupta Efficiency KGE, Bias Ratio BR, correlation R, variability ratio γ, and root mean square error normalized by reference precipitation mean NRMSE) of 1-day lead GFS forecasts and different satellite products, during the wet periods, at different watersheds.

10.2.7 How is the Performance of GFS Affected if the Reference Product is Changed from IMERG Final to CHIRPS?

We acknowledge that the reference dataset used in our evaluation (i.e., IMERG Final) has its own estimation errors. We conducted additional assessment to evaluate the performance of GFS using CHIRPS rainfall products as reference. Table 10.2 shows the performance statistics of GFS for different lead times, using IMERG FINAL and CHIRPS, separately, rainfall products as reference. The overall magnitude of GFS performance (as well as the variability of the performance across watersheds) is similar when either rainfall product is used as a reference. Both reference products reveal that the GFS forecast has low skill in the two highly mountainous watersheds (Hartbeespoort and Loskop), but relatively better skill in the larger watershed, Massingir. The GFS forecasts show slightly lower performance when CHIRPS is used as a reference compared to having IMERG Final as reference. Therefore, our results show that the overall performance of GFS remains the same if the reference product were to be changed from IMERG Final to CHIRPS, indicating the robustness of IMERG Final as reference product.

Table 10.2. Comparison of the performance statistics of the 1-day lead GFS forecast when using IMERG Final (CHIRPS) as reference, for each of the dam watershed.

Watershed	Correlation	Bias Ratio	KGE
Gaborone	0.53 (0.35)	0.88 (0.95)	0.51 (0.35)
Mokolo	0.51 (0.54)	1.09 (0.91)	0.48 (0.53)
Hartbeespoort	0.53 (0.46)	1.40 (1.45)	0.39 (0.28)
Loskop	0.49 (0.39)	1.29 (1.30)	0.38 (0.27)
Massingir	0.66 (0.54)	1.04 (1.09)	0.66 (0.52)

* information obtained from the Global Reservoir and Dam Database (Lehner et al. 2011) and Food and Agriculture Organization of the United Nations (FAO)'s Global Information System on Water and Agriculture (AQUASTAT).
** Calculated from HydroSEHDS (Lehner et al. 2008).

10.3 Conclusions

The accuracy of medium-range (1-day to 15-day lead time) precipitation forecasts available from the Global Forecast System (GFS) were evaluated in the Limpopo River Basin, focusing on the watersheds of five major dams in the basin: Gaborone (watershed area of 5,353 km2, mean elevation of 1,183 m.a.s.l.), Mokolo (5,195 km2, 1,305 m.a.s.l.), Hartbeespoort (4,991 km2, 1,492 m.a.s.l.), Loskop (15,162 km2, 1,544 m.a.s.l.), and Massinger (68,779 km2, 910 m.a.s.l.). The evaluation was made using two satellite-gauge rainfall products as reference, IMERG Final and CHIRPS. The performance of watershed-average 1-day lead GFS forecasts as measured through the KGE statistics varies within the range 0.27 to 0.66. The KGE values are relatively low at 0.39 (0.28) and 0.38 (0.27) for IMERG Final (CHIRPS) reference rainfall product, for the two high-elevation watersheds, Hartbeespoort and Loskop, respectively. The KGE is relatively high at 0.66 (0.52) for the large and relatively low-elevation watershed (Massinger). The KGE for the medium-elevation and small-sized watersheds, Gaborone and Mokolo watersheds, ranges from 0.35 to 0.53.

The performance of GFS decreases with increasing lead time, and this is attributed to the decreasing ability of the forecast to capture the temporal fluctuation of daily rainfall as portrayed by IMERG Final and CHIRPS. The performance of GFS also decreases with increasing the temporal aggregation scale. The KGE values fall sharply at about 5-day lead time. IMERG Early captures the temporal fluctuation of daily rainfall, and therefore, it can be considered as a data source to post-process GFS forecasts in order to improve the accuracy of GFS forecasts, especially at long lead times. Finally, we point out that this study is based on one year of data, which does not represent a full range of climate conditions, however, the key findings of this study could be considered for further improvement of the accuracy of GFS products.

CHAPTER 11

Evaluation of Medium-Range Forecasts (GFS) over Orange

11.1 Study Region

The Orange River Basin (Figure. 11.1), with a drainage area of about 950,000 km2, stretches over four countries – Lesotho, South Africa, Botswana, and Namibia (Cambray et al. 1986; Benade 1988). The riparian countries' contribution to the water supply in the Orange River is as follows (Lange et al. 2007): Lesotho (34%), South Africa (64%), Namibia (2%), and Botswana (1%). The contribution in terms of drainage area land is as follows (Lange et al. 2007): Lesotho (3%), South Africa (64%), Namibia (25%), and Botswana (4%). We note that, Lesotho, the most upstream country that falls entirely within the basin and occupies only 3.4% of the basin area, contributes 34% of the water supply in the Orange River. The Orange River rises in the east in the steep mountains of Lesotho, flows through the savannah grasslands of the central plateau, and to the desert conditions in the west, near the Atlantic Ocean (Bohensky et al. 2004). The Orange River plays an important role in the South African economy by providing water for irrigation and hydroelectric power. Several dams were built in the Orange River system and as a result the Orange River is the most regulated system in South Africa (Benade 1988). For this study we selected four major dams on account of their size and spatial representativeness (Table 1). The Hardap Dam, located in the western part of the watershed, is the largest dam Namibia primarily used for irrigation and flood control. The Kalkfontein, Erfenis, and Vaal dams are located in the eastern parts of the watershed. The Kalkfontein Dam and Erfenis dams in South Africa are primarily used for irrigation, while the former is also used for flood protection. The Vaal Dam in South Africa is primarily used for domestic and industrial water supply, as well as for flood protection. The watershed areas of these dams vary within the range of about 6,000 km2 to about 49,000 km2.

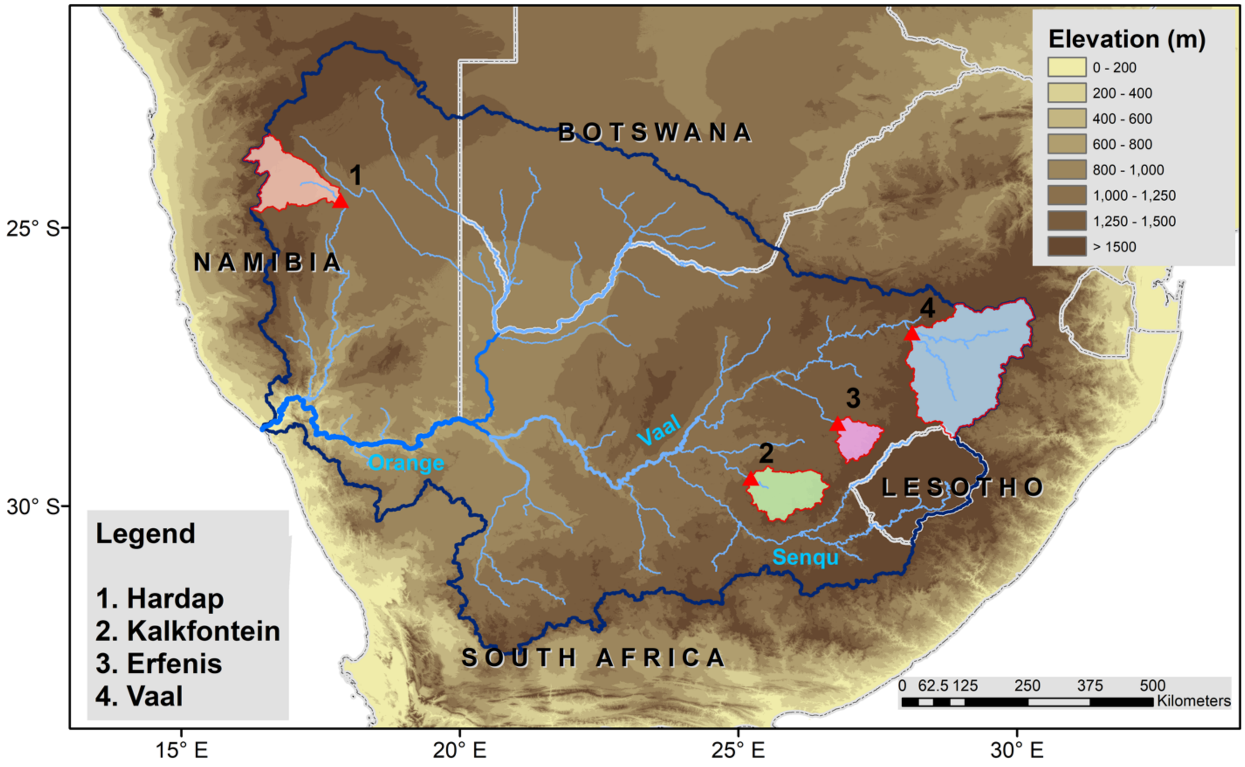


Figure 11.1. Map of Orange River Basin, with the location of selected dams/reservoirs: (1) Hardap, (2) Kalkfontein, (3) Erfenis, and (4) Vaal, and the drainage basins defined by the dam locations (see the colored areas between dams).

For this study, we selected four dams located in different parts of the basin (Figure 11.1, Table 11.1). The selected dams have drainage areas ranging from 6,182 km^2 (Erfenis Dam) to 48,791 km^2 (Vaal Dam). The Vaal Dam in South Africa has the largest storage capacity in the basin at 2536 Mm^3, and is a dam built mainly for flood control. The Hardap Dam in Namibia, with a storage capacity of 294.6 Mm^3, is a multi-purpose dam built for irrigation and flood control. The Kalkfontein Dam has a capacity of 258.3 Mm^3, and mainly serves for irrigation and flood control. The Erfenis Dam in South Africa has a storage capacity of 212.3 Mm^3, and supplies water to the agriculture region.

Table 11.1 Major reservoir dams in the Limpopo River Basin

Dams	Country	Operational Since*	Capacity (million m^3)*	Purpose*			Area of Drainage Basins (km^2)**	Elevation of Drainage Basin (m)**
				Irrigation	Flood Control	Hydroelectricity		
Hardap	Namibia	1962	294.6	x	x		16128	1421
Kalkfontein	South Africa	1977	258.3	x	x		13601	1423
Erfenis	South Africa	1960	212.3	x			6182	1484
Vaal	South Africa	1985	2536		x		48791	1675

* information obtained from the Global Reservoir and Dam Database (Lehner et al. 2011) and Food and Agriculture Organization of the United Nations (FAO)'s Global Information System on Water and Agriculture (AQUASTAT).
** Calculated from HydroSEHDS (Lehner et al. 2008).

11.2 Results and Discussion

11.2.1 Annual Spatial Variability of Rainfall

Figure 11.2 displays the spatial map of annual (15 June 2019 to 15 June 2020) rainfall derived from the 1-day lead GFS forecast and satellite precipitation products. The reference product, IMERG Final, shows rainfall gradient in the west to east direction, with the western part characterized by very low rainfall (< 100 mm yr^{-1}) and the eastern part by relatively higher rainfall (> 100 mm yr^{-1}). Validated against IMERG Final, the 1-day lead GFS captures the overall gradient, but tends to underestimate rainfall in the dry (western) parts of the watershed and overestimate rainfall in the wet (eastern) parts. The correlation between the spatial distribution of GFS forecast and IMERG Final, at annual timescale, is only 0.17, indicating that GFS has difficulty capturing the spatial distribution of precipitation.

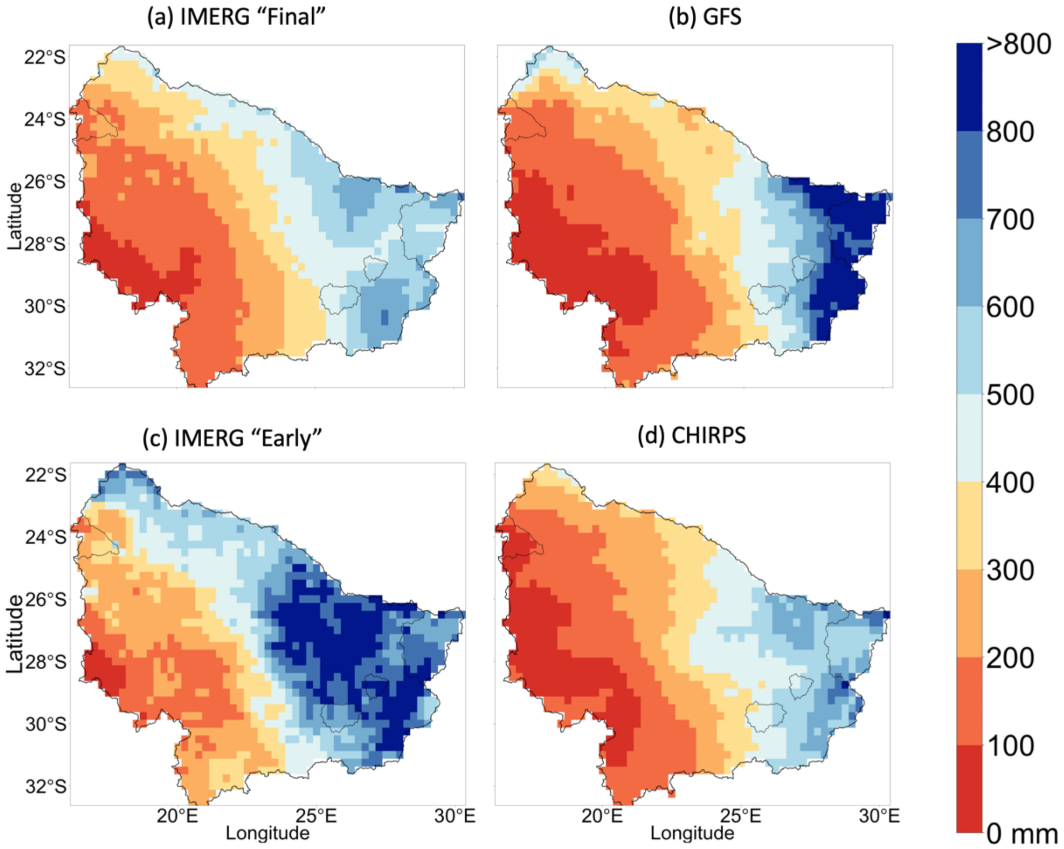


Figure 11.2. Spatial map of annual rainfall (in mm), for the period 15 June 2019 to 15 June 2020, derived from IMERG Final, GFS (1-day lead time), IMERG Early, and CHIRPS, for watersheds of five dams: (1) Hardap, (2) Kalkfontein, (3) Erfenis, and (4) Vaal.

11.2.2 Monthly Cycle of Rainfall

Figure 11.3 presents the time series of monthly rainfall derived from the 1-day lead GFS and satellite precipitation products. According to IMERG Final, the region is characterized by a distinct wet season, November to April, and a dry season, May to October. The GFS monthly time series agrees with IMERG Final, but tends to overestimate in the wettest parts of the watershed (i.e. Vaal dam).

11.2.3 Annual Rainfall

Here, we aggregate the 1-day lead GFS forecasts and satellite rainfall estimates to annual time scale (Figure 11.4). According to IMERG Final, the annual watershed-average rainfall is 197 mm (Hardap), 485 mm (Kalkfontein), 519 mm (Erfenis), and 557 mm (Vaal). The 1-day lead GFS captures well the annual rainfall in the western part (Hardap) and some parts (Kalkfontein) of the watershed, but overestimates over Erfenis (by 18%) and Vaal (by 54%).

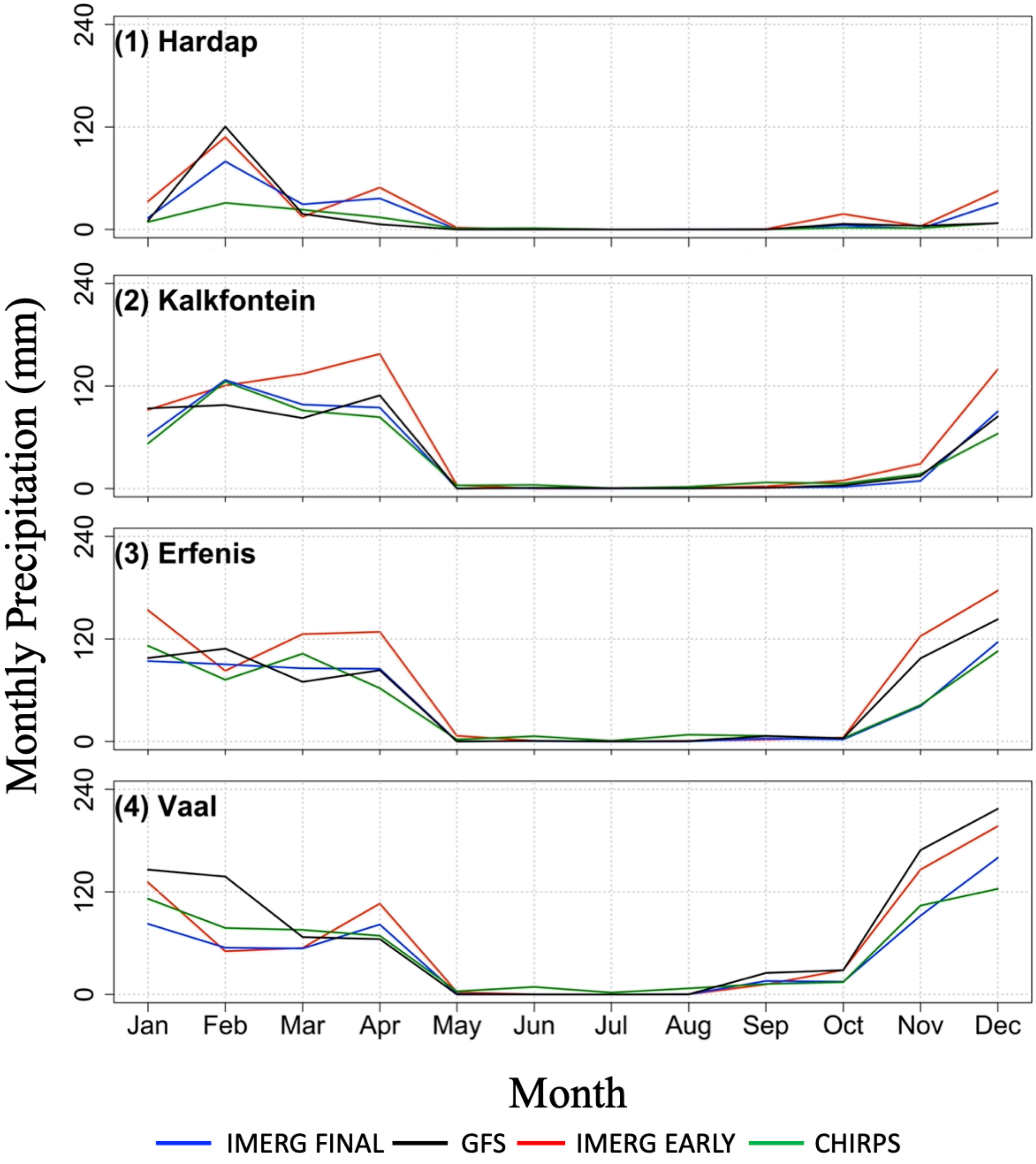


Figure 11.3 Monthly time series of sub-basin averaged precipitation (mm), for the period 15 June 2019 to 15 June 2020, derived from IMERG Final, GFS (1-day lead time), IMERG Early, and CHIRPS, for watersheds of four dams: (1) Hardap, (2) Kalkfontein, (3) Erfenis, and (4)Vaal.

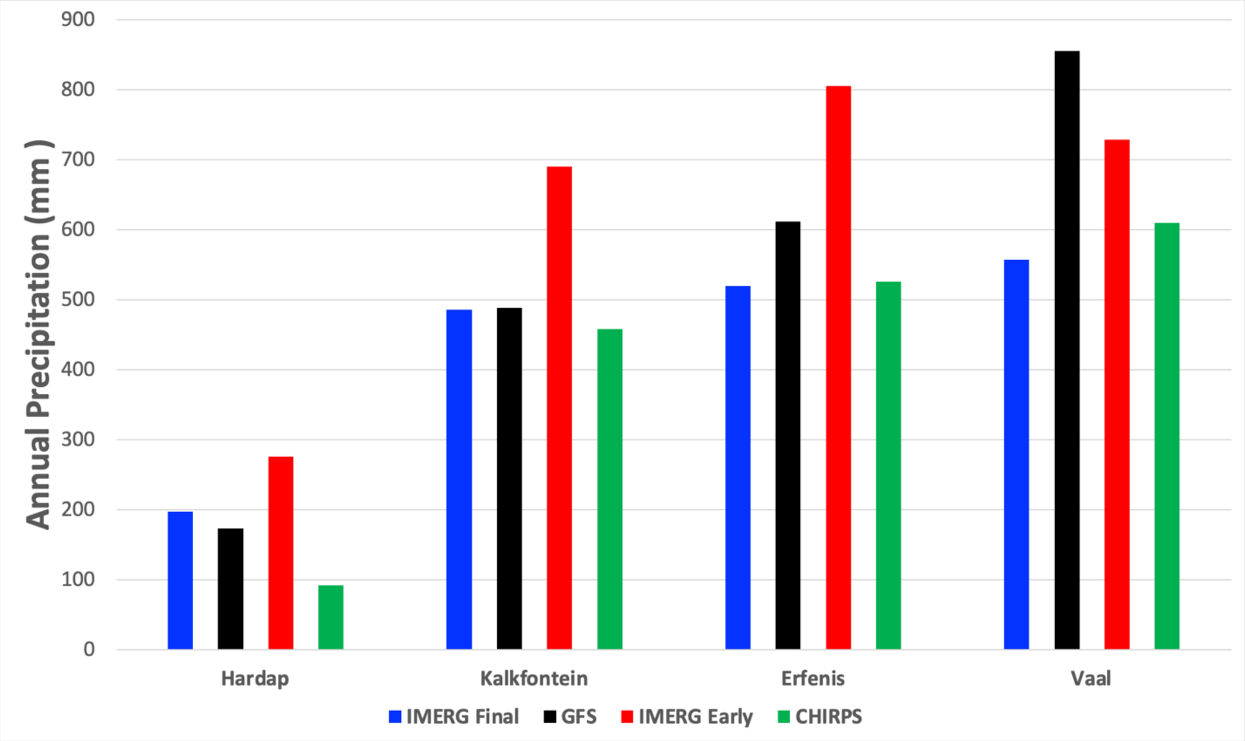


Figure 11.4. Sub-basin averaged annual precipitation (mm) for the period, 15 June 2019 to 15 June 2020, for each of the Orange's sub-basin, derived from the 1-day lead GFS forecast and different satellite precipitation products.

11.2.4 Daily Time Series

Figure 11.5 presents the daily time series of watershed-averaged rainfall, derived from 1-day lead GFS and satellite products. According to IMERG Final, the coefficient of variation (CV) of daily rainfall varies in the range 1.66 to 2.13 in the eastern (and relatively wet) parts of the watershed, while it is 2.40 in the western and dry parts. The GFS produces similar CV in the eastern part, but overestimates CV in the western part, and misses observed peaks and produces false alarms in the western part.

11.2.5 KGE Statistics

Figure 11.6 presents the performance statistics of various precipitation products (with respect to the IMERG Final estimates), in terms of Kling-Gupta Efficiency (KGE), Bias Ratio (BR), correlation (R), variability ratio (□), and root mean square error normalized by reference precipitation mean (NRMSE). The KGE of 1-day lead GFS forecast varies in the range 0.24 to 65. The KGE is poor (0.24 and 0.37) in Hardeep (where the GFS overestimates the coefficient of variation) and Vaal (where the GFS has high overestimation bias). The KGE of GFS is intermediate (0.65 and 0.52) in Kalkfontein and Erfenis watersheds. The breakdown of the KGE scores (BR, R, and γ) reveals the key factors contributing to the KGE estimates. The bias of GFS is worse for Vaal (large overestimation). The correlation coefficient between GFS and IMERG Final is very low for Hardap. The variability ratio for GFS is worse for Hardap. The normalized RMSE is very high at around 300% of the mean for Hardap, and it is around 180% for all three dam watersheds. The KGE breakdown indicates that the required developments to improve the representation of daily precipitation in GFS forecasts is improving the bias in the case of Vaal watershed and improving the temporal and spatial fluctuation in the case of Hardap.

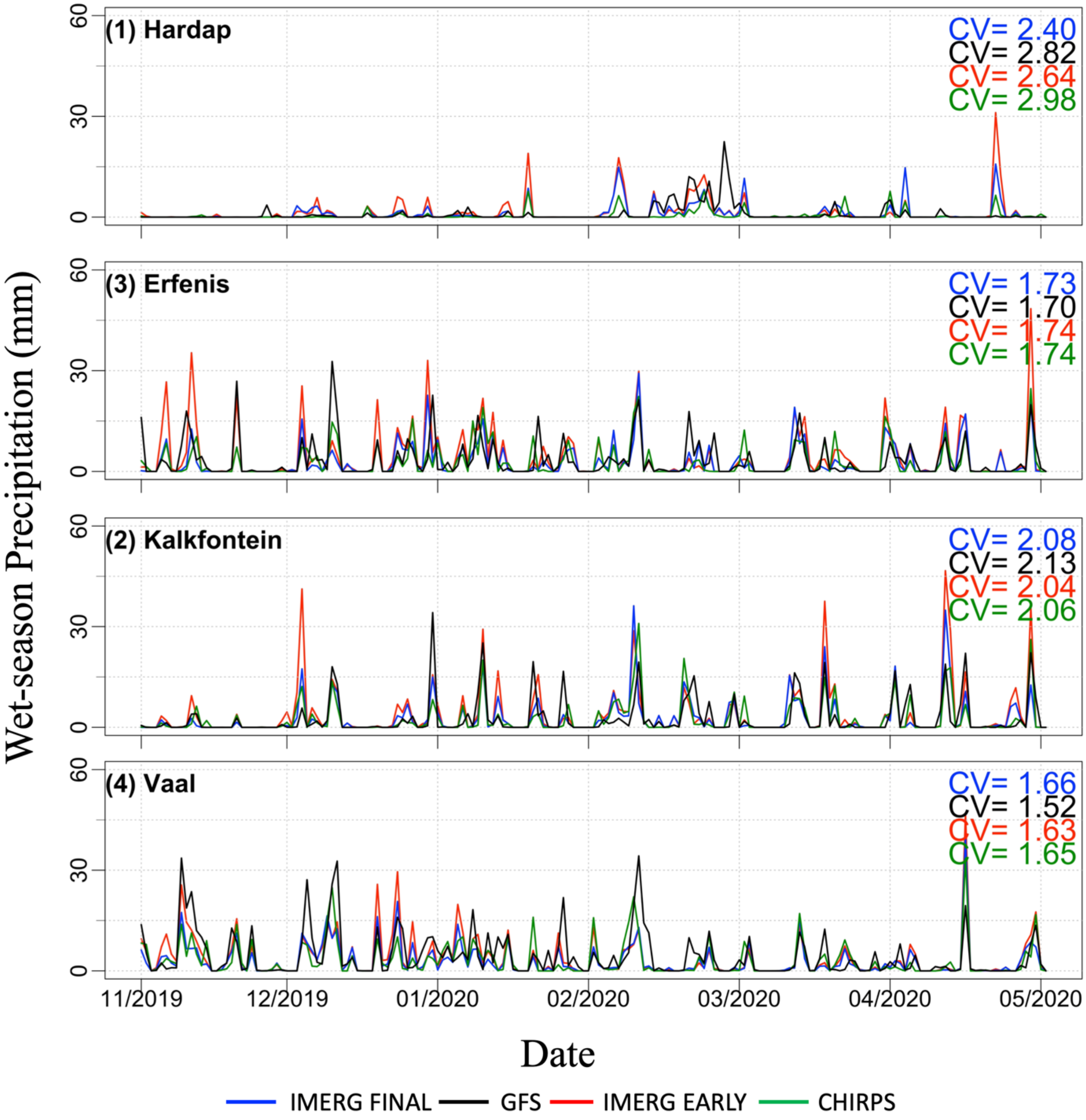


Figure 11.5. Time series of sub-basin averaged precipitation total (mm) for the wet period (November – April), for each of the dam watersheds, as derived from various precipitation products. The Figure also shows the coefficient of variation (CV) as a measure of temporal variation.

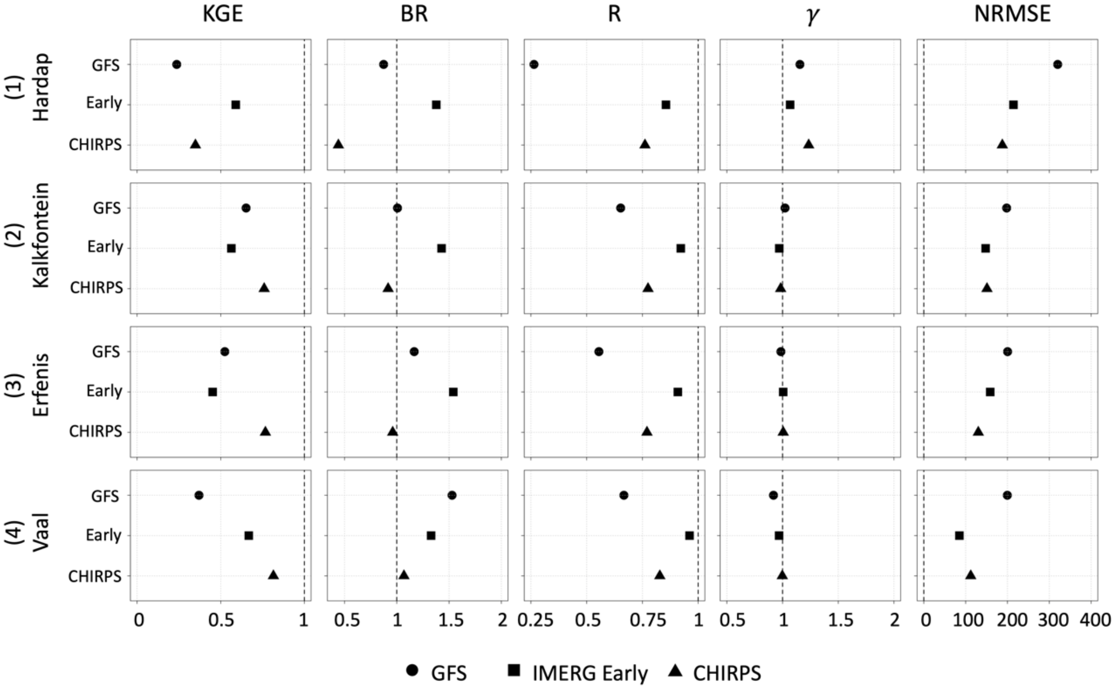


Figure 11.6 Summary of performance statistics (Kling-Gupta Efficiency KGE, Bias Ratio BR, correlation R, variability ratio γ, and root mean square error normalized by reference precipitation mean NRMSE) of 1-day lead GFS forecasts and different satellite products, during the wet periods, at different watersheds.

11.2.6 Dependence of Forecast Performance on Precipitation Rate

Figure 11.7 presents the scatterplot of 1-day lead GFS forecasts against daily IMERG Final rain rates. Over Hardap (a dry region), the GFS has very low correlation with IMERG Final (R = 0.27) and underestimate all rain rates exceeding 5 mm day^{-1}. Over Vaal (the wettest region), the GFS has moderate correlation (R = 0.67), but high false alarm and large overestimation at light and medium rain rates. Over Erfenis and Kalkfontein, the GFS forecasts have large scatter around IMERG Final estimates.

11.2.7 Effect of Lead Time on Forecast Performance

Figure 11.8 presents the Kling-Gupta Efficiency (KGE) and its components for 5-day total rainfall forecast of GFS for three different lead time periods, 1-5 day, 6-10 day, and 10-15 day. The KGE values are higher at short lead times (1 to 5 day) than at longer lead times (6 to 10 day, and 11 to 15 day). The KGE values at short lead times are higher by about 0.50 compared to the longer lead times. The KGE values fall sharply at 5-day lead time, and KGE does not change much beyond a lead time of 5 days. The breakdown of the KGE scores (BR, R and γ) reveals the key factors contributing to the differences in KGE estimates across lead times. The main differences in KGE statistics across the lead times can be primarily attributed to the correlation statistics: the correlation between the 6-10 day (or the 11-15 day) and the IMERG Final is under 0. 45 in all watersheds, but jumps to 0.45 to 0.75 for the correlation between 1-5 day and IMERG Final, in all watersheds. Therefore, in order to improve the accuracy of GFS forecasts at longer lead times, the temporal fluctuation of GFS forecasts needs to be improved further to match that for IMERG Final.

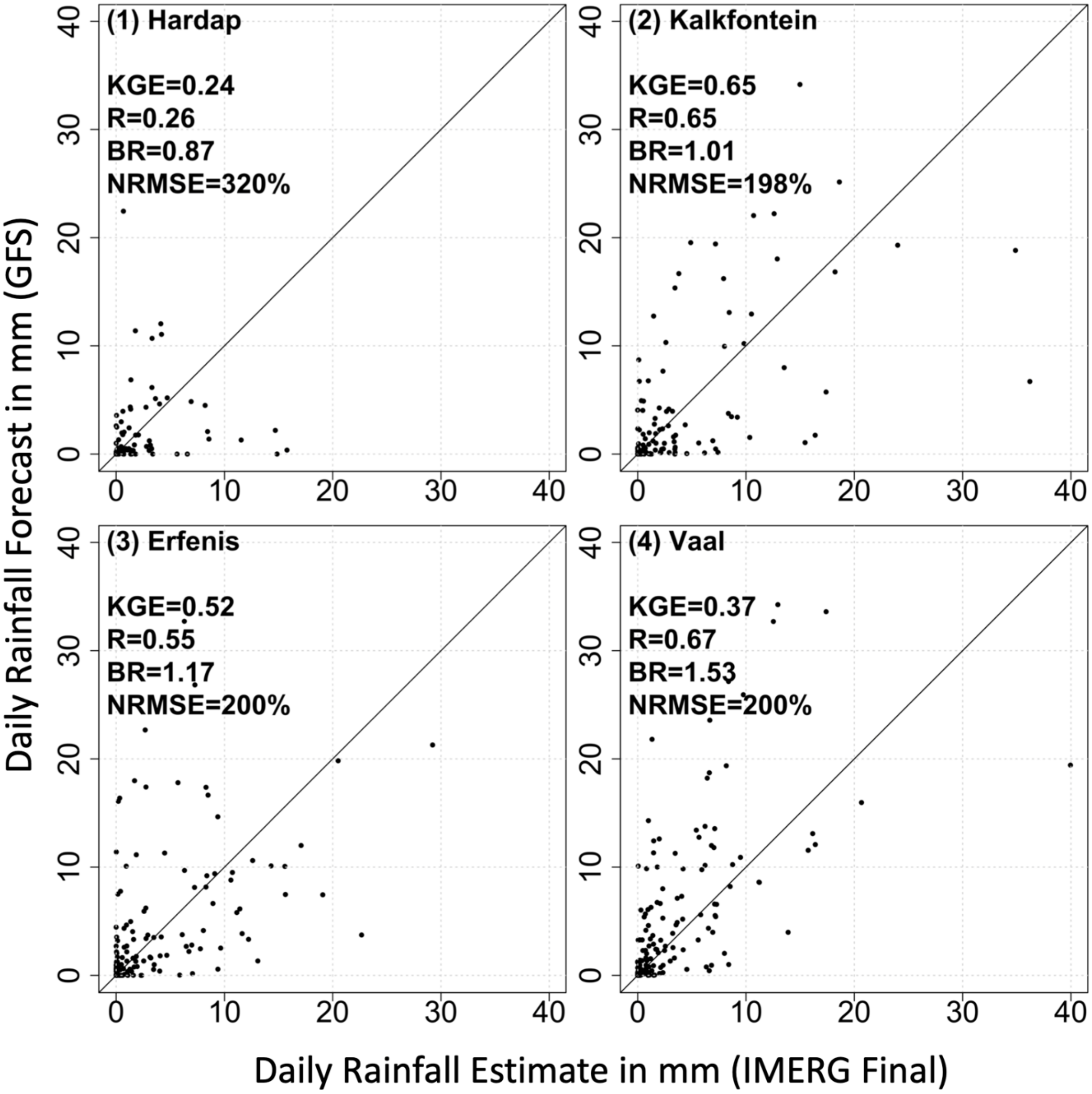


Figure 11.7. Scatterplot of watershed-averaged daily precipitation forecast obtained from the 1-day lead GFS forecast against corresponding values from IMERG Final, for five dam watersheds.

11.2.8 Effect of Temporal Aggregation Scale on Forecast Performance

Figure 11.9 presents the scatterplot of watershed-averaged 15-day accumulated rainfall obtained from the GFS forecasts against IMERG Final estimates, for each dam watershed. The 15-day total forecast is obtained by adding multiple lead-time daily forecasts: 1-day lead, 2-day lead, and all the way up to 15-day lead. Over all the watersheds, the 15-day accumulated rainfall shows inferior performance compared to the daily GFS forecast (KGE deteriorated from 0.24 at daily to 0.18 at 15-day accumulation for Hardap, from 0.65 to 0.49 for Kalkfontein, from 0.52 to 0.43 for Erfenis, and from 0.37 to 0.05 for Vaal). We point out that aggregation over longer timescales (from daily to 15-day total) is subject to two different error natures: effect of averaging on forecast errors, and effect of lead time on forecast errors. We anticipate that the effects of lead time on errors to dominate as averaging over longer time scale is expected to reduce random errors.

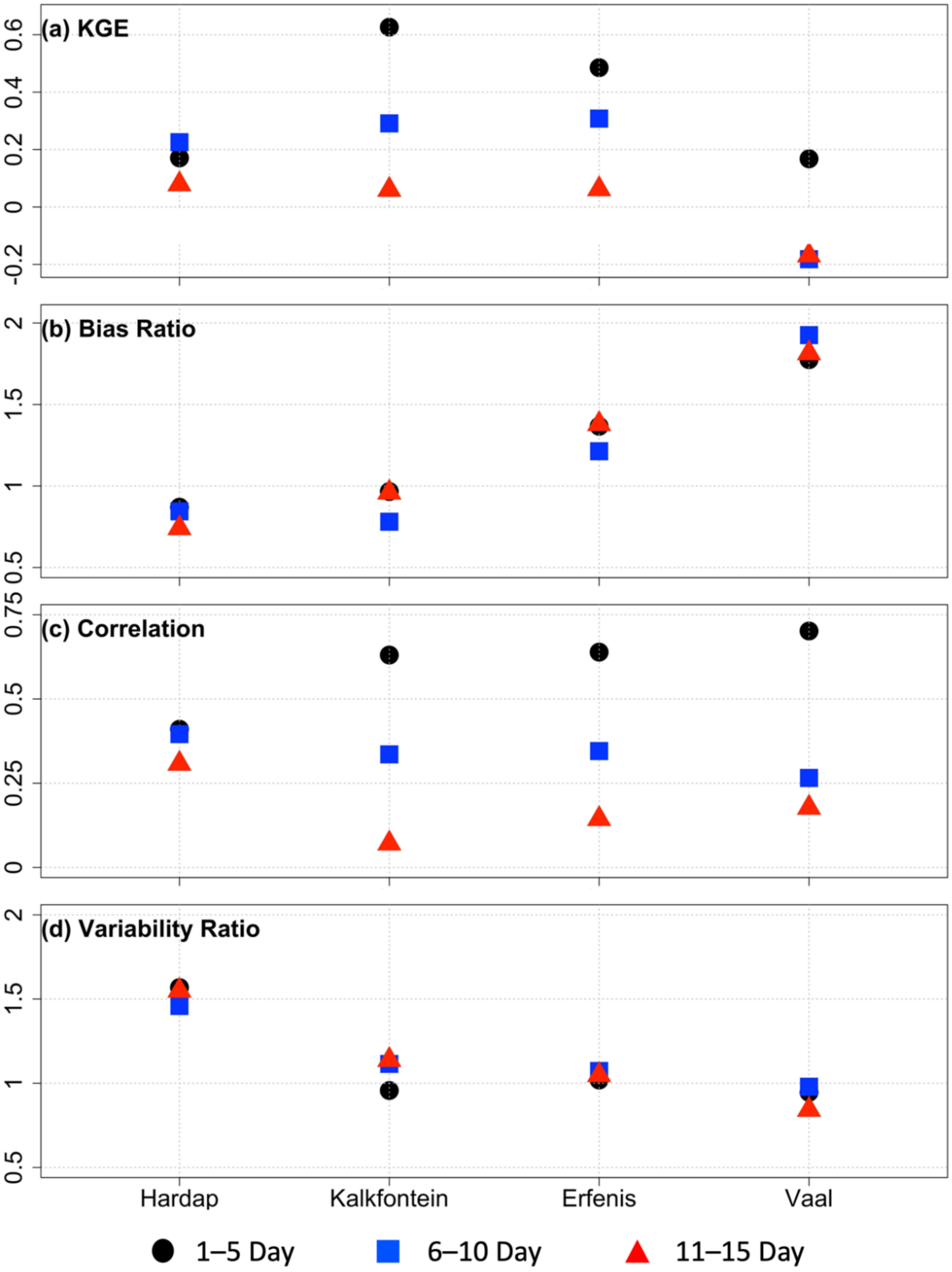


Figure 11.8. Kling-Gupta Efficiency (KGE) and its components for 5-day total rainfall forecast of GFS for three different lead time periods, 1-5 day, 5-10 day, and 10-15 day, for each dam watershed.

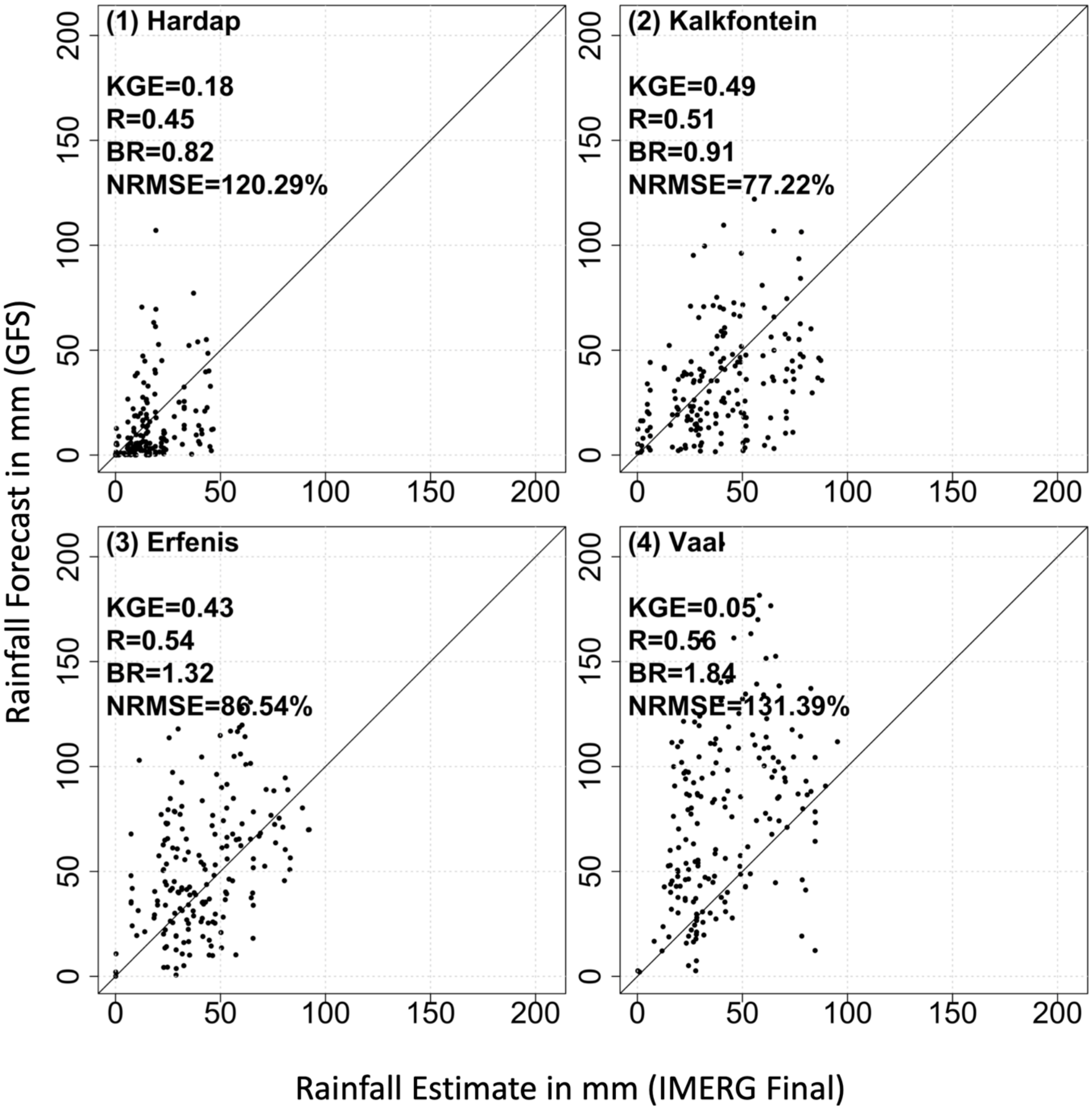


Figure 11.9. Scatterplot of watershed-averaged daily precipitation forecast obtained from the 1-day lead GFS forecast against corresponding values from IMERG Final, for five dam watersheds.

11.2.9 Comparison of the Performances of IMERG Early and GFS

The poor skills of GFS at Hardap and Vaal watersheds and the intermediate skills of GFS at Kalkfontein and Erfenis watersheds suggest the need for exploring ways of improving the GFS forecast accuracy. One such method is post-processing of GFS forecasts using rainfall estimates that have relatively better accuracy and are available in near-real time. Satellite-only products, such as IMERG Early, are available in near-real time. In this section, we explore how the performance of IMERG Early compare with the performance of GFS. The spatial distribution of IMERG Early has better correlation with IMERG Final (R = 0.91), compared to GFS, but suffers from large overestimation bias over the entire eastern (and wet) part of the watershed (Figure 11.2). In terms of monthly rainfall values (Figure. 11.3), IMERG Early outperforms GFS over the two watersheds where GFS has relatively poor performance (Hardap and Vaal), but GFS outperforms IMERG Early over the other two watersheds where GFS has intermediate skill (Kalkfontein and Erfenis). In terms of annual rainfall values (Figure. 11.4), IMERG Early has overestimation bias at all sites, with GFS outperforming IMERG Early at all watersheds except for Vaal watershed where IMERG Early has lower overestimation bias then IMERG Early. As far as the daily time series of rainfall is concerned, IMERG Early has good agreement with IMERG Final (Figure. 11.5). In terms of KGE, IMERG Early outperforms GFS over the two watersheds (Hardap and Vaal) where GFS shows poor skills, but GFS outperforms IMERG Early over the other two watersheds (Kalkfontein and Erfenis) where GFS shows intermediate skills. The reason for the relatively low performance of IMERG Early over Kalkfontein and Erfenis can be attributed to the high bias in IMERG Early.

11.2.10 Performance of GFS if the Reference Product is Changed from IMERG Final to CHIRPS

We acknowledge that the reference dataset used in our evaluation (i.e., IMERG Final) has its own estimation errors. We conducted additional assessment to evaluate the performance of GFS using CHIRPS rainfall products as reference. Table 11.2 shows the performance statistics of GFS for different lead times, using IMERG Final and CHIRPS, separately, rainfall products as reference. The overall magnitude of GFS performance (as well as the variability of the performance across watersheds) is similar when either rainfall product is used as a reference. Both reference products reveal that the GFS forecast has low skill in the Hardap and Vaal watersheds, but relatively better skill in the Kalkfontein and Erfenis watersheds. The GFS performances using IMERG Final and CHIRPS are similar for the three watersheds located in the eastern (and wet) parts of the watershed. However, the GFS performance with CHIRPS as reference is found to be worse than with IMERG Final as reference for the western (and dry) watershed, Hardap, which has very low skill even with IMERG Final as reference. Therefore, our results show that the overall performance of GFS remains the same if the reference product were to be changed from IMERG Final to CHIRPS, indicating the robustness of IMERG Final as reference product.

Table 11.2 Performance statistics of daily GFS forecast over the four major dam watersheds of the Orange River Basin. for various lead times (1-day, 5-day, 10-day, and 15-day) using IMERG Final (CHIRPS) rainfall products as reference, in terms of correlation, bias ratio, and NRMSE.

Lead time of GFS forecast	Correlation	Bias Ratio	KGE	NRMSE (%)
		Hardap		
1-day	0.25 (0.21)	0.86 (1.95)	0.21 (-0.24)	270 (575)
5-day	0.06 (0.20)	0.54 (1.22)	-0.21 (0.11)	311 (508)
10-day	0.15 (0.15)	0.94 (2.14)	0.12 (-0.42)	336 (656)
15-day	0.13 (0.06)	0.79 (1.78)	0.10 (-0.23)	300 (558)
		Kalkfontein		
1-day	0.64 (0.68)	1.01 (1.10)	0.64 (0.66)	180 (179)
5-day	0.38 (0.22)	0.81 (0.89)	0.35 (0.22)	214 (246)
10-day	-0.02 (-0.03)	0.78 (0.85)	-0.04 (-0.04)	267 (276)
15-day	0.05 (0.05)	0.89 (0.98)	0.04 (0.04)	273 (285)
		Erfenis		
1-day	0.53 (0.55)	1.17 (1.22)	0.50 (0.50)	182 (185)
5-day	0.25 (0.26)	1.46 (1.52)	0.12 (0.09)	260 (269)
10-day	-0.02 (-0.03)	1.25 (1.30)	-0.04 (-0.07)	274 (285)
15-day	-0.09 (-0.08)	1.39 (1.45)	-0.16 (-0.17)	287 (294)
		Vaal		
1-day	0.65 (0.64)	1.53 (1.42)	0.36 (0.44)	185 (173)
5-day	0.45 (0.45)	2.05 (1.91)	-0.21 (-0.09)	259 (242)
10-day	0.12 (0.11)	1.93 (1.80)	-0.28 (-0.20)	324 (306)
15-day	0.01 (0.07)	1.99 (1.85)	-0.44 (0.31)	288 (265)

11.3 Conclusions

We evaluated the accuracy of medium-range (1-day to 15-day lead time) forecasts available from the Global Forecast System (GFS) in the Orange River Basin, focusing on some of its major dams, namely, Hardap, Kalkfontein, Erfenis, and Vaal. While Hardap is located in the dry (and western) parts of the watersheds, the remaining three dams are located in the wet (and eastern) parts of the watershed. The evaluation is done using the satellite-gauge merged rainfall product IMERG Final as reference, and additional analysis is conducted by changing the reference product to CHIRPS. Additionally, the performance of GFS is compared to the performance of the near-real time, satellite-only rainfall product, IMERG Early, to see if IMERG Early has the potential to be used in post-processing of GFS.

Our results show the following:

- The 1-day lead GFS forecast captures the west-east gradient in rainfall, but tends to overestimate in the eastern parts of the region. The 1-day lead GFS overestimates over the watershed located in the far eastern part of the watershed (Vaal) by about 53%, while the bias over the other three watersheds is in the range -13% to 17%.
- The correlation between the daily watershed-averaged 1-day lead GFS forecast and IMERG Final is relatively high in the eastern (and wet) watersheds (Kalkfontein, Erfenis, and Vaal), with correlation ranging from 0.55 to 0.67, but is low in the western (and dry) watershed (Hardap) with correlation of 0.26.
- The KGE of 1-day lead GFS forecast is low (KGE = 0.24) in the western (Hardap) due to the low correlation and (KGE = 0.37) in the far eastern (Vaal) due to its high bias, whereas it is relatively higher (KGE = 0.52 and 0.65) in the remaining two watersheds (Erfenis and Kalkfontein, respectively).
- The KGE forecast accuracy decreases as the lead time increases. The KGE values are relatively high for 1-day lead time, and drop sharply about 5-day lead time, beyond which they remain about the same.
- Aggregating the forecasts from daily to longer timescales (say 15-day accumulation) decreases the performance of GFS forecasts. This indicates that the reduction in random error expected by aggregating forecasts over long time scale is dominated by the increasing error that comes with introducing the lead time of forecasts.

- While IMERG Early has better performance than GFS over the two watersheds where GFS gives low KGE values (Hardap and Vaal), IMERG Early has lower performance than GFS over the watersheds where GFS gives higher KGE values ((Erfenis and Kalkfontein) due to the high bias of IMERG Early estimates over the latter two watersheds.

Therefore, we conclude that the accuracy of GFS forecasts in the Orange River Basin depends on the location of the watershed as well as the lead time and aggregation temporal scale. The accuracy is relatively low over the dry (western) and far eastern (wet) parts of the watershed, while it is relatively higher in the eastern parts of the watershed. The high bias of IMERG Early in the eastern watershed implies that bias removal of IMERG Early needs to be conducted before its usage in any potential post-processing of GFS forecasts based on IMERG Early datasets.

CHAPTER 12

Conclusion

The main objective of this dissertation was to improve the understanding of the performance of recently updated short-range and medium-range precipitation forecasts: (1) the assessment of HRRR, a short-range precipitation forecasts product for early extreme events warning in US, and (2) the assessment of GFS, a medium-range precipitation forecasts product for reservoir management in Africa. The motivation of this dissertation arises from the need to evaluate the skills of short-range and medium-range forecasts so they can be used in confidence in applications, such as early flood warning and reservoir management, and to identify the strengths and drawbacks of the current forecast quality in order to provide valuable insights to algorithm developers and product users. In this section, a summary of the conclusion drawn from the research described in Chapter 3 through Chapter 11 is presented.

12.1 Conclusion from the Evaluation of Short-Range Forecasts

This study evaluated the accuracy of short-range (lead times ranging from 1 h to 18 h) forecasts, for five extreme events in the United States which covered two frontal storms and two hurricanes: the September 21-23, 2016, frontal storms in Iowa, (2) the April 28-May 1, 2017, frontal storms in the Southern Midwestern US, (3) the August 25-31, 2017, Hurricane Harvey storms in Texas, (4) the September 13-17, 2018, Hurricane Florence storms, and (5) the September 4-6, 2019, Hurricane Dorian storms in the Carolinas. The basis of the investigation was the HRRR operational forecasts, which are used as input into the National Weather Service' National Water Model (NWM). Evaluation of the forecasts was carried out by comparison with high-quality and independent rainfall observational products known as the gauge-corrected

Multi-Radar/Multi-Sensor (MRMS-GC). The main results are summarized as follows.

- There was a good agreement between area-averaged forecasts and observations, on an hourly scale. However, the forecasts were mostly biased. The forecasts tend to overestimate rainfall for both hurricanes. However, the forecasts tend to underestimate the frontal storm in Iowa but produced almost unbiased estimates for the Southern Midwestern US.
- The forecasts were able to capture the spatial pattern of hurricanes, albeit with overestimation. However, the forecasts produced too many, localized, high-rain intensities for the frontal storms. In addition, the forecasts have difficulty locating the single supercell for the frontal storm in Iowa.
- With regard to the effect of lead time, the 1 h lead forecast had lower accuracy compared to the other lead-time forecasts. For lead times ranging from 2 h to 18 h, there was not much systematic difference in accuracy among the various lead-time forecasts.
- The bias estimates for the small spatial scale varied quite a lot, mostly within the range of -100% to +100%, indicating that the bias estimates obtained at large scale (hundreds of km grids) are not applicable to bias estimates at smaller spatial scales, and vice versa. The bias did not also show significant reduction as the rainfall averaging grid increases from 2 km × 2 km all the way to 32 km × 32 km.

In conclusion, the results of our investigation show that the forecasts captured well the temporal variability of observed precipitation, indicating that the HRRR forecasts provide relatively reliable forecasts. In comparison, the forecasts have better accuracy for predicting hurricanes compared to frontal storms, particularly those frontal storms with single super cells.

Our results also show that the 1-h lead forecasts showed generally lower accuracy than the other lead-time forecasts.

Finally, we point out that, although the selected storm cases are interesting from meteorological perspective, they are small in number. Thus, the findings of this study can only provide a first insight into the accuracy of HRRR forecasts for extreme precipitation. Additional analysis involving more storm cases and mechanistic approaches needs to be carried out in order to generalize the results.

12.2 Conclusion from the Evaluation of Medium-Range Forecasts

In Chapter 4 through Chapter 11, the performance of Global Forecast System (GFS) was evaluated over eight transboundary river basins (Nile, Niger, Volta, Senegal, Congo, Zambezi, Limpopo and Orange River Basin) in Africa using satellite-gauge merged precipitation observations, namely IMERG and CHIRPS. A couple of sub-basins in each river basin are defined here on the basis of major reservoir dam locations. The main results are summarized as follows.

- The performance of GFS varies in different regions. GFS forecasts have “intermediate” or “poor” skills over West Africa, East Africa and Southern Africa while GFS has “very poor” KGE scores over Central Africa (see also Figure. 12.1).
- The GFS has large overestimation bias in most cases, except for the watersheds in the dry regions (see also Figure. 12.2).

- In terms of capturing the temporal dynamics of observed watershed-averaged precipitation, the majority of the watersheds show “intermediate” or “poor” correlation for GFS (see also Figure. 12.3).
- The GFS forecast accuracy decreases as the lead time increases, but the rate of decrement depends on the region.
- Aggregating the forecasts at temporal scales (1-day to 15-day) may increase or decrease the performance of GFS forecasts, depending on the region (see Figure. 12.4).
- IMERG Early has complimentary advantages over GFS in most basins (see Figure. 12.5). Given that IMERG Early outperforms GFS, we recommend testing the suitability of IMERG Early to serve as input into post-processing of GFS in order to improve the accuracy of GFS forecasts.

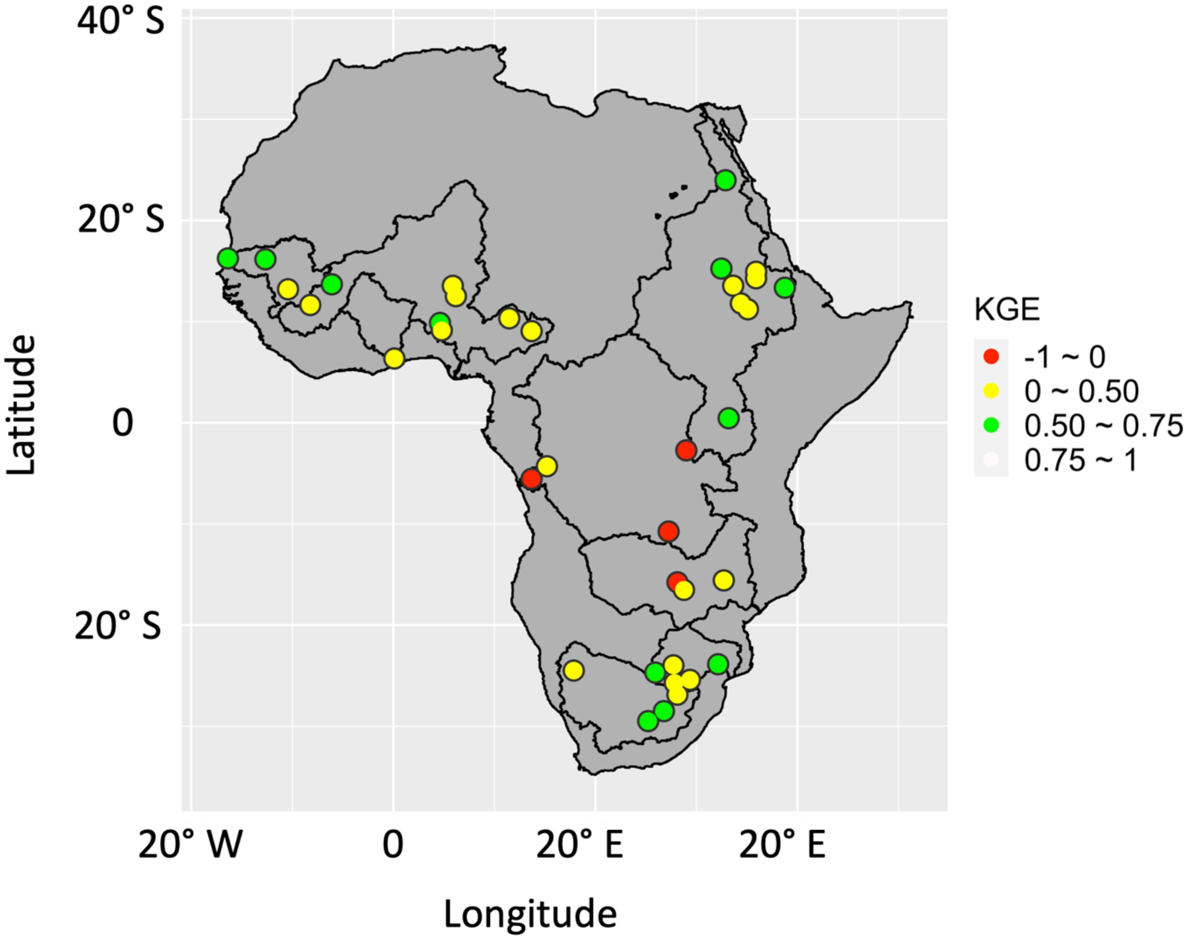


Figure 12.1 Summary of Kling-Gupta Efficiency KGE of 1-day lead GFS forecasts for different dam watersheds in Africa.

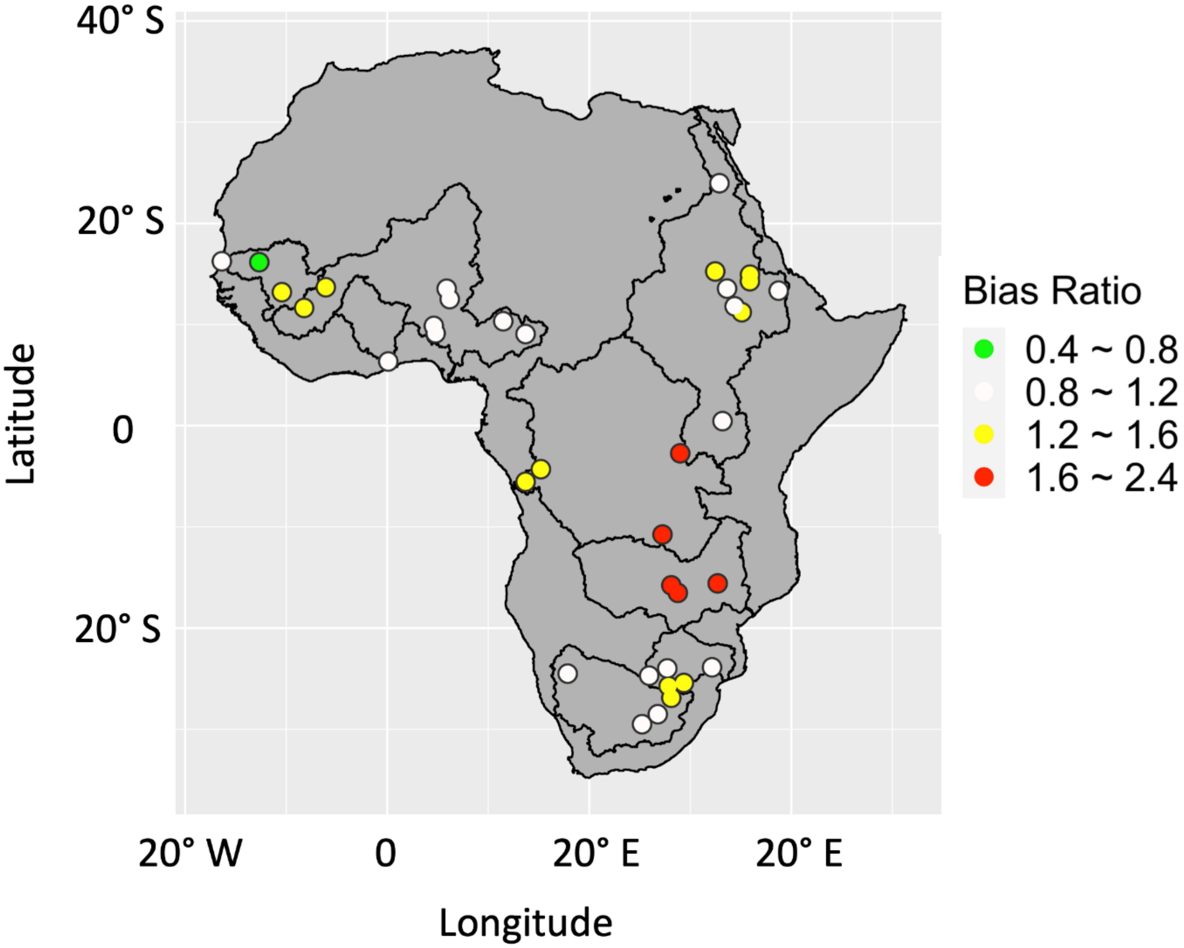


Figure 12.2 Summary of bias ratio of 1-day lead GFS forecasts for different dam watersheds in Africa.

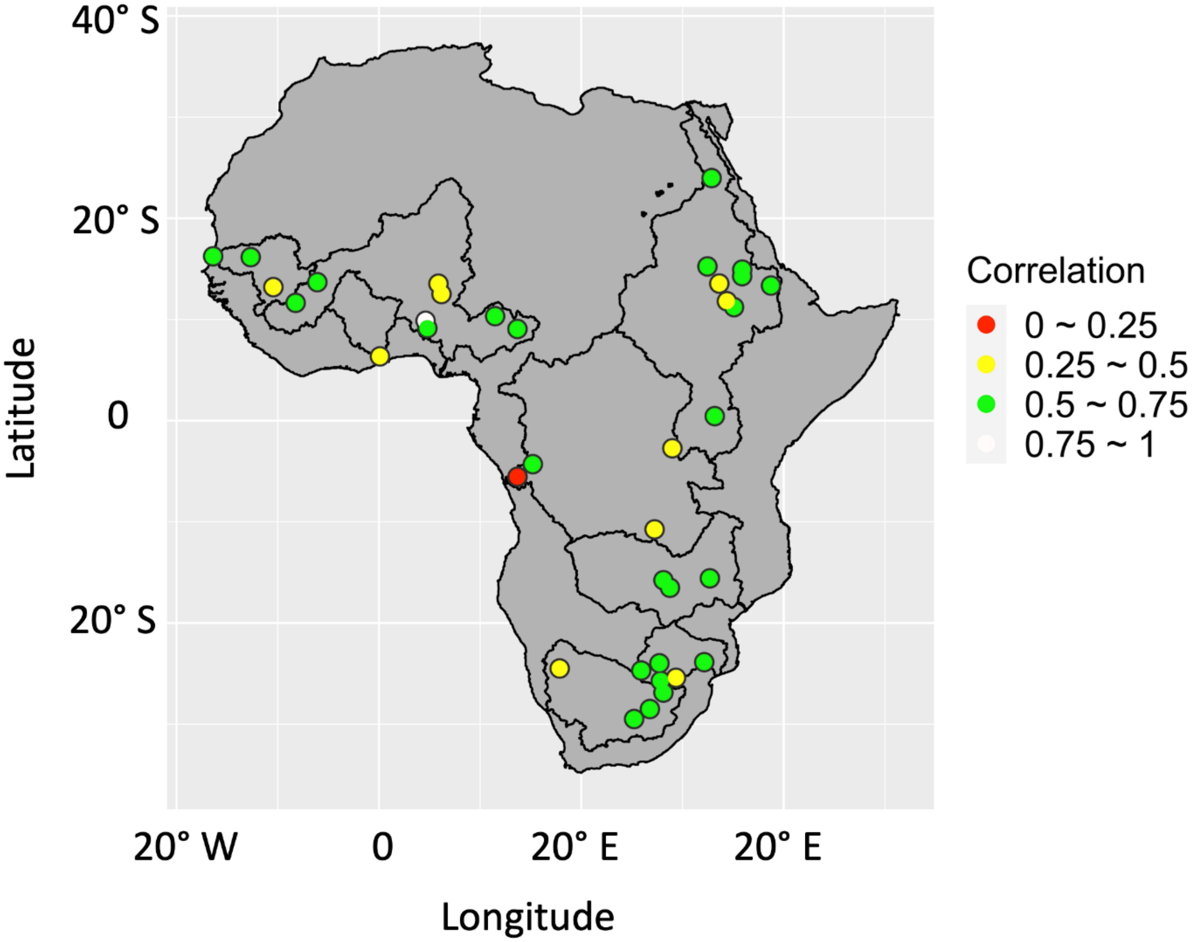


Figure 12.3 Summary of correlation of 1-day lead GFS forecasts for different dam watersheds in Africa.

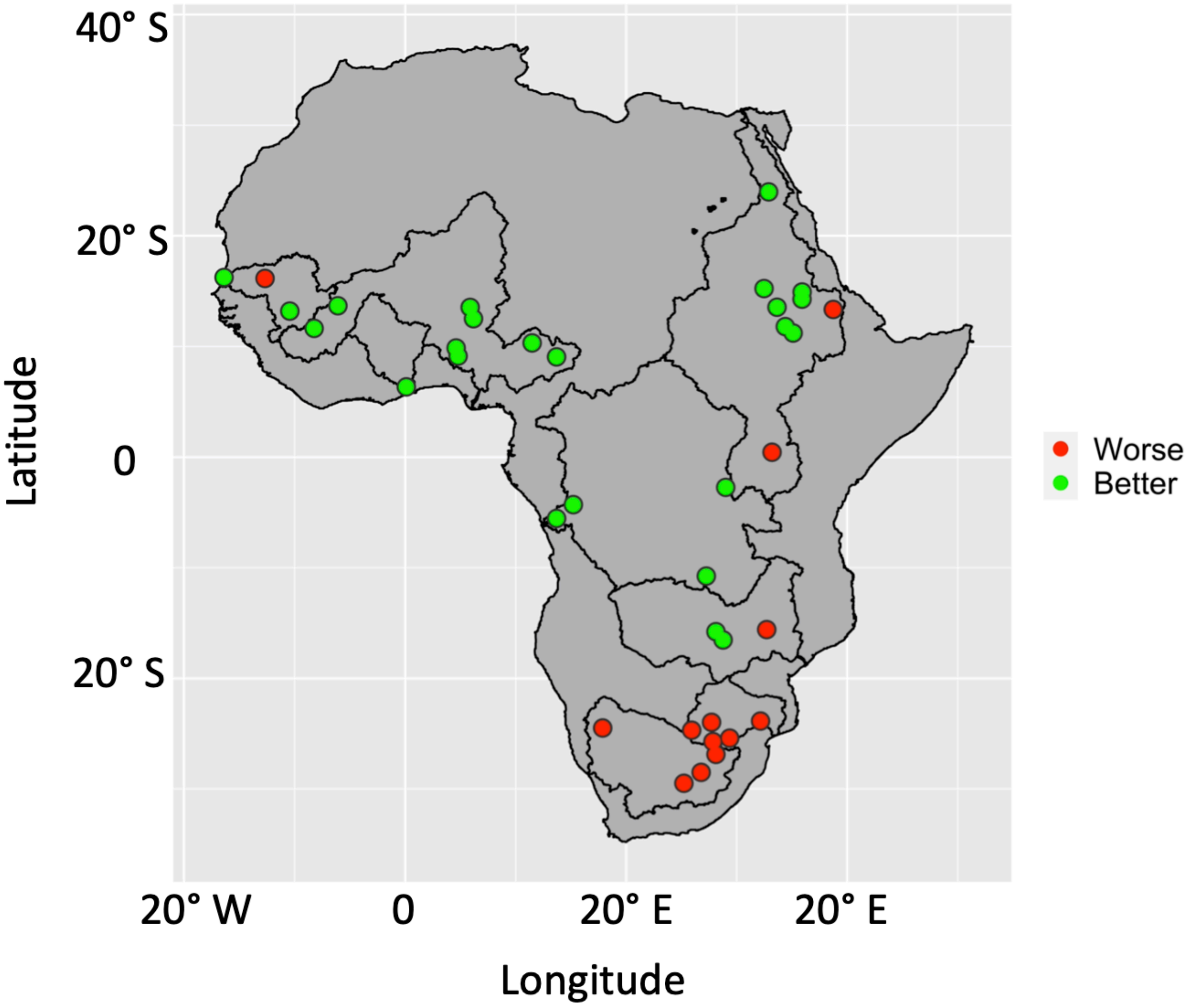


Figure 12.4 Comparison of KGE scores of 1-day lead GFS and 15-day accumulated GFS

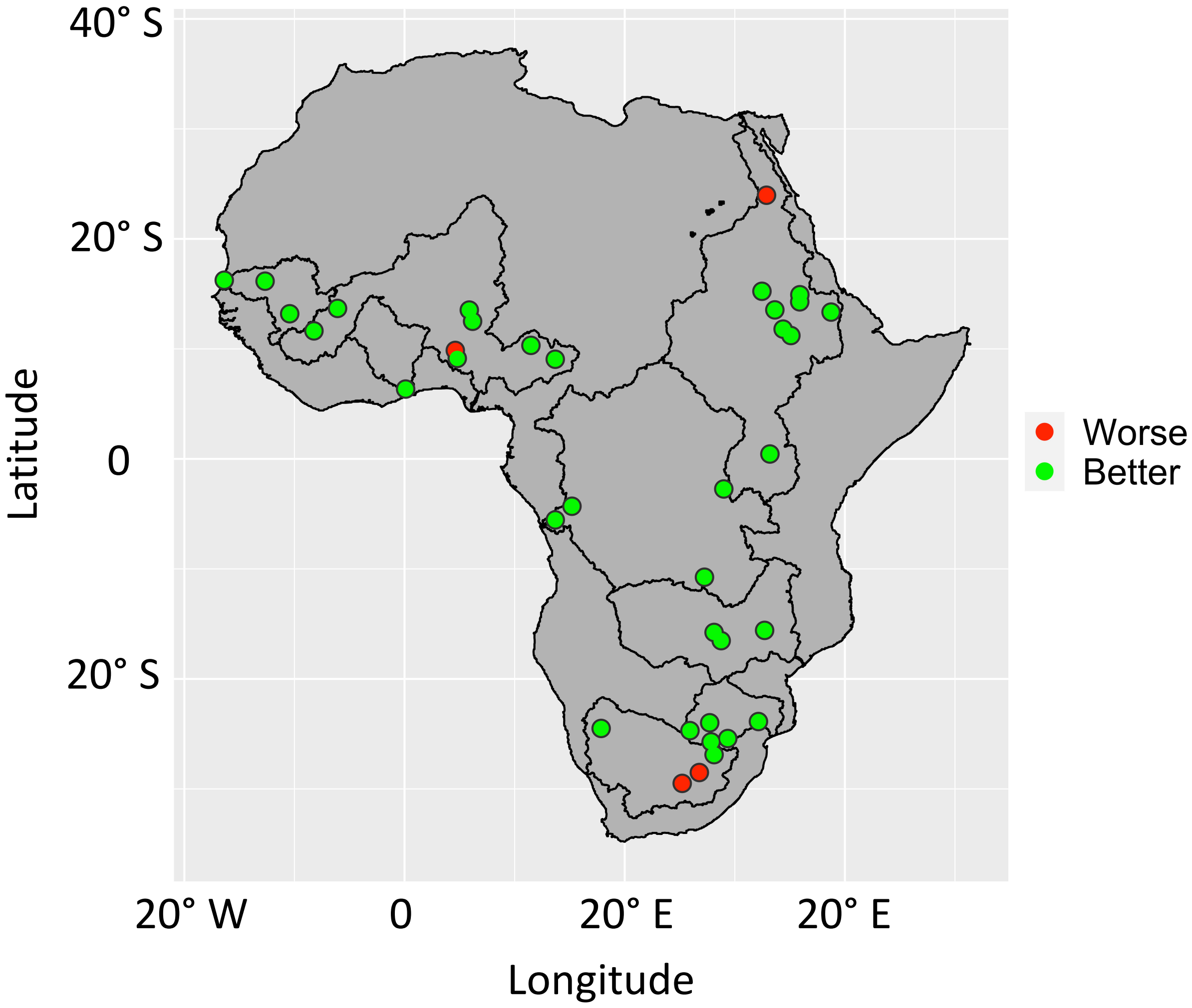


Figure 12.5 Comparison of KGE scores of 1-day lead GFS and IMERG Early

Appendix A

This appendix provides supporting information for Chapter 2.

This chapter has been accepted in the Journal of Hydrometeorology. © American Meteorological Society. Used with permission.

Yue, H., Gebremichael, M. and Nourani, V., 2021. Evaluation of Global Forecast System (GFS) Medium-Range Precipitation Forecasts in the Nile River Basin. *Journal of Hydrometeorology*. accepted. doi: 10.1175/JHM-D-21-0110.1

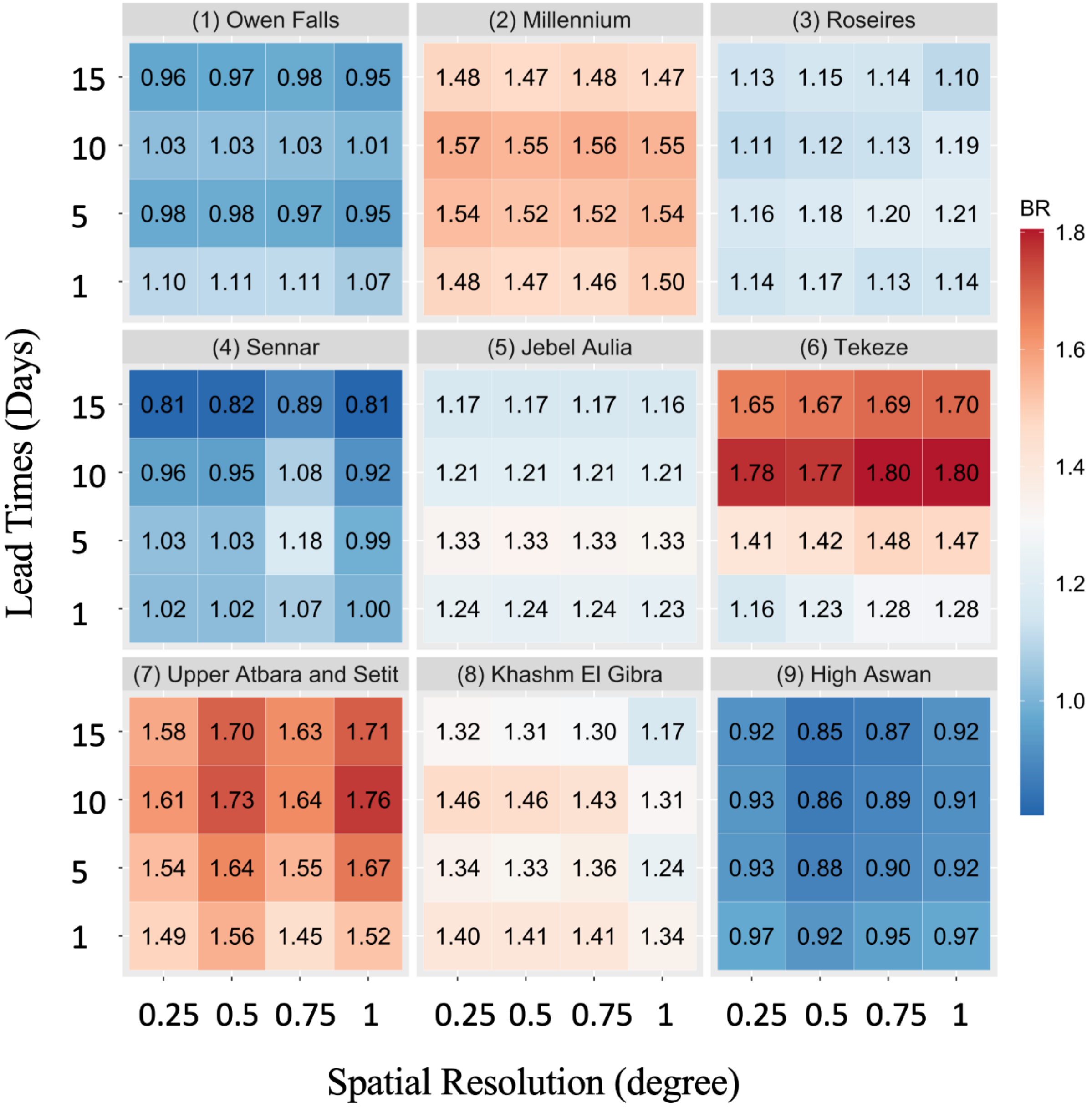


Figure A1. Bias ratio (BR) for daily precipitation of GFS compared to IMERG Final, for GFS lead times of 1-day, 5-day, 10-day, and 15-day, and at different spatial scales, for different Nile sub-basins.

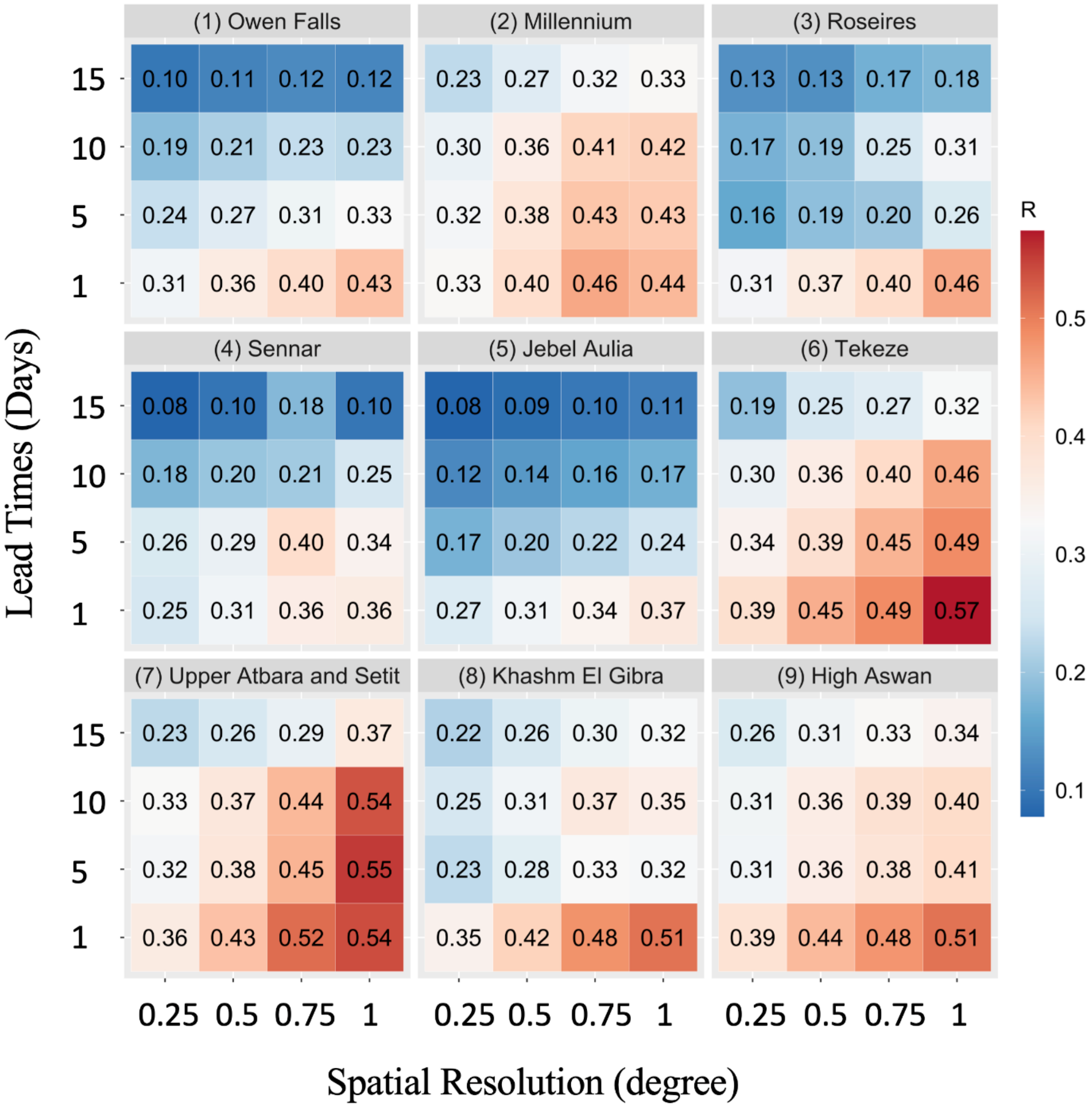


Figure A2. Correlation (R) for daily precipitation of GFS compared to IMERG Final, for GFS lead times of 1-day, 5-day, 10-day, and 15-day, and at different spatial scales, for different Nile sub-basins.

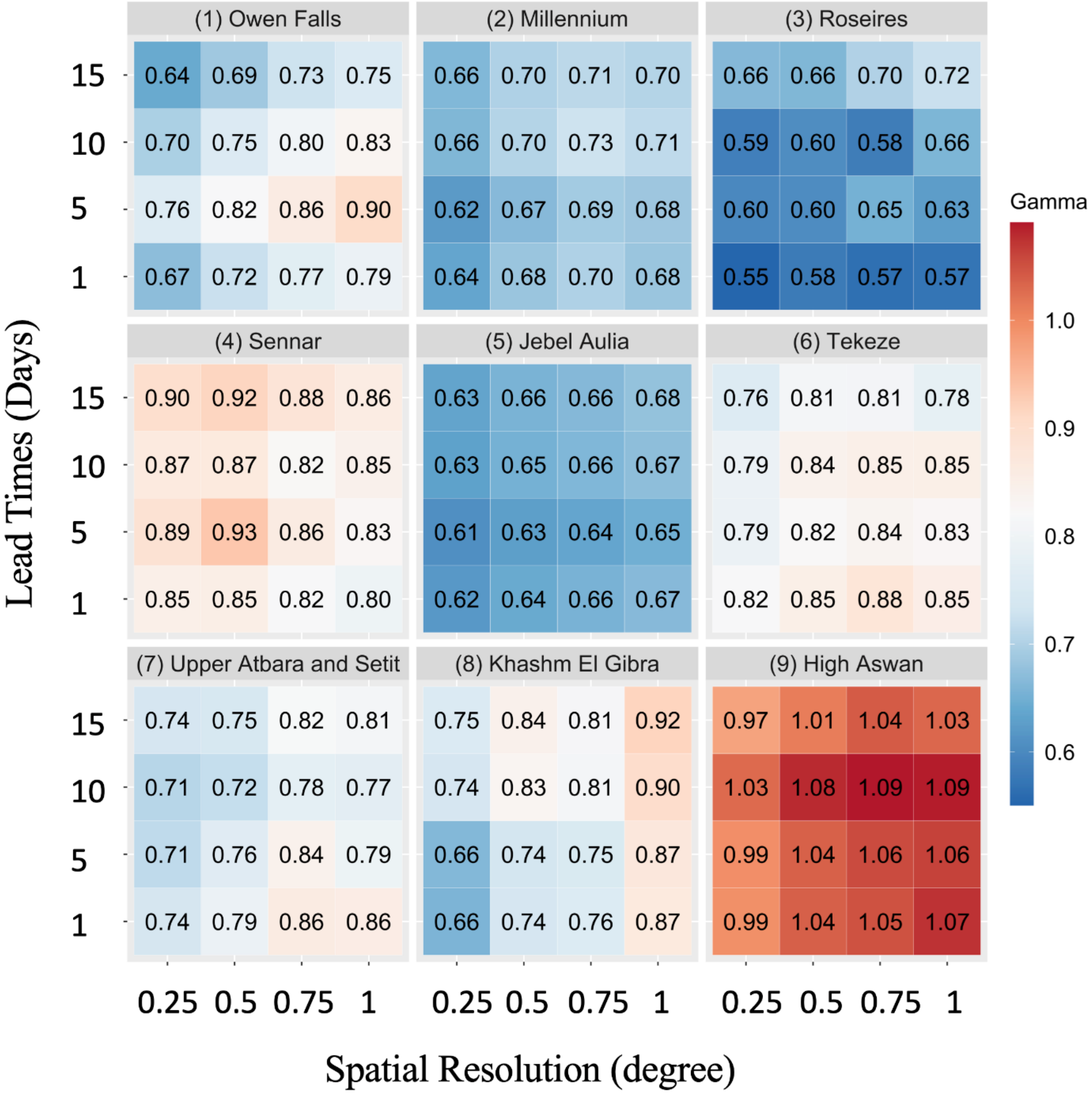


Figure A3. Variability ratio for daily precipitation of GFS compared to IMERG Final, for GFS lead times of 1-day, 5-day, 10-day, and 15-day, and at different spatial scales, for different Nile sub-basins.

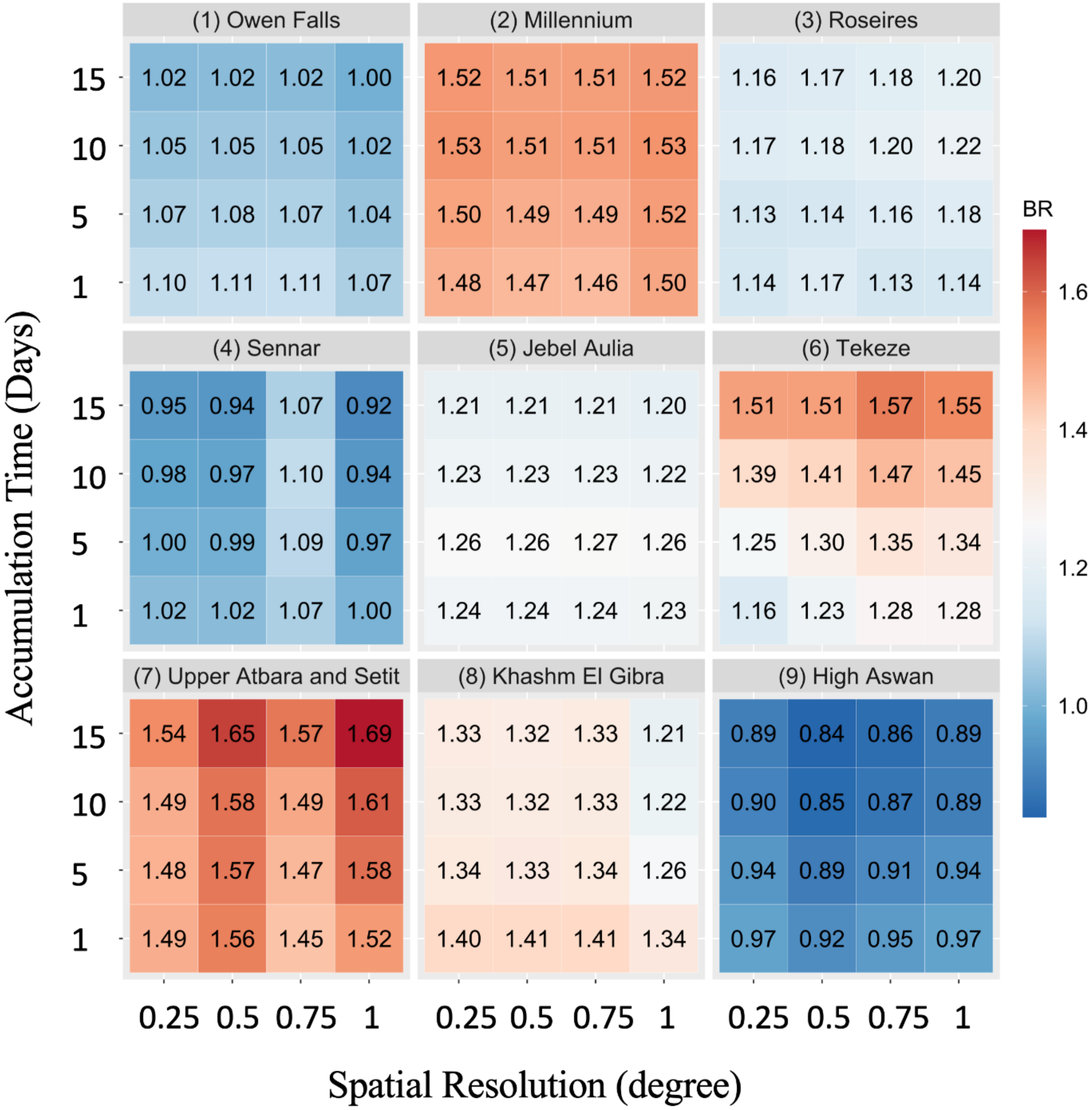


Figure A4. Bias Ratio (BR) for GFS precipitation compared to IMERG Final, for accumulation time scales of 1-day, 5-day, 10-day, and 15-day, and at different spatial scales, for each Nile sub-basin.

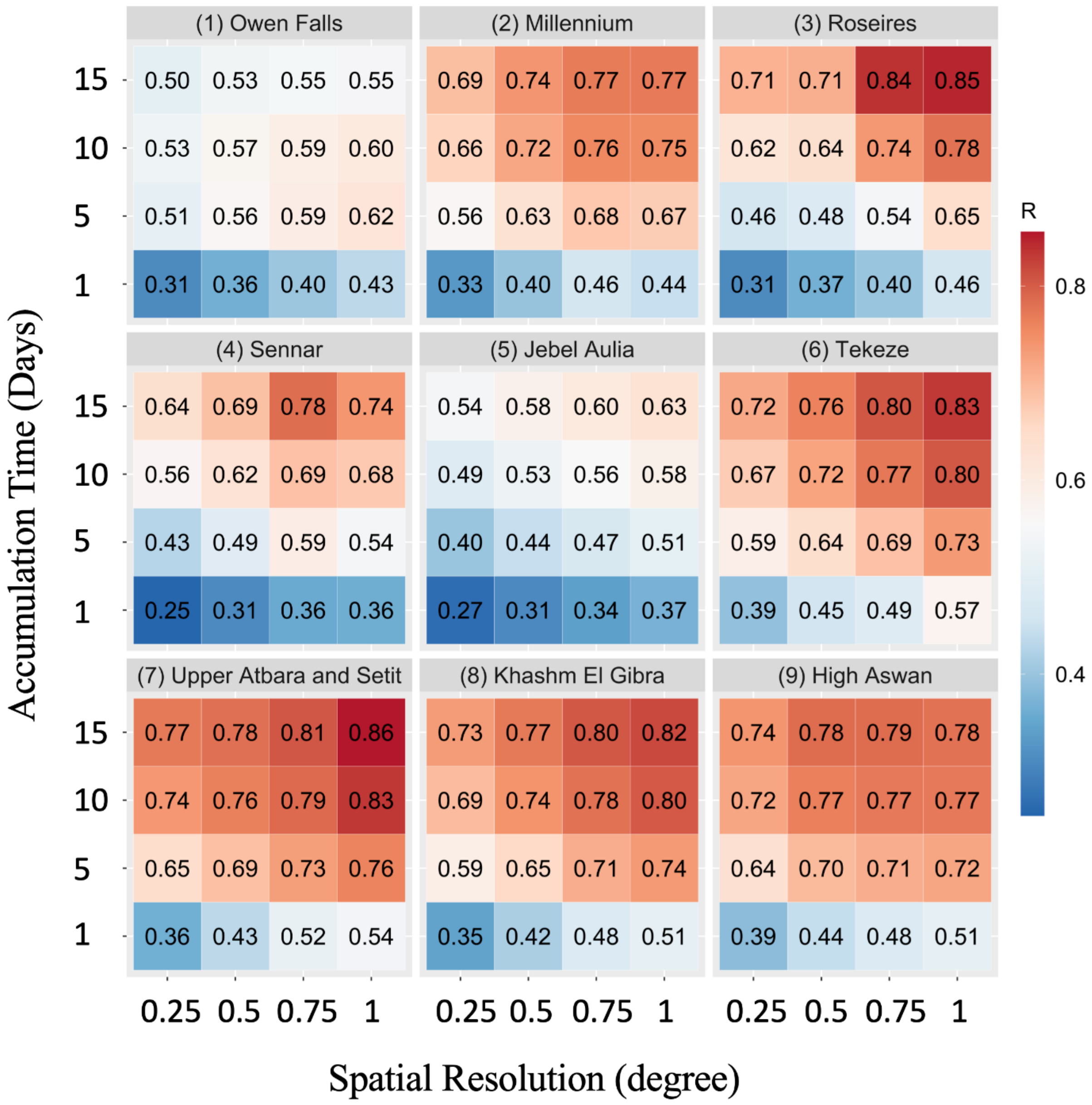


Figure A5. Correlation for GFS precipitation compared to IMERG Final, for accumulation time scales of 1-day, 5-day, 10-day, and 15-day, and at different spatial scales, for each Nile sub-basin.

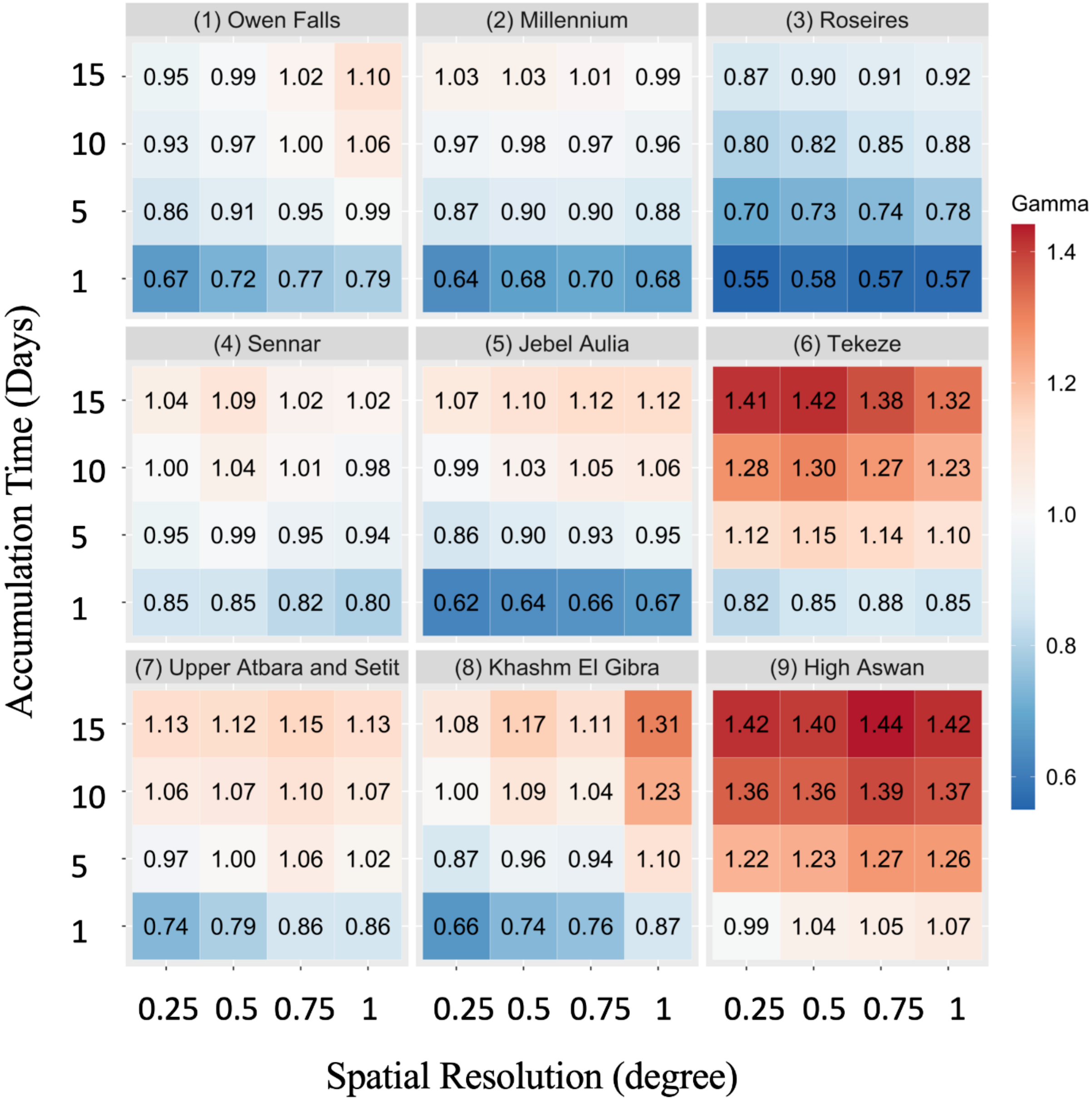


Figure A6. Variability ratio for GFS precipitation compared to IMERG Final, for accumulation time scales of 1-day, 5-day, 10-day, and 15-day, and at different spatial scales, for each Nile sub-basin.

Bibliography

Abdalla, S., Janssen, P. A., Balmaseda, M. A., Mogensen, K., Bidlot, J. R., & Keeley, S., 2013 December. ECMWF Experience In Global Monitoring And Validation Of Radar Altimetry Products. In ESA Living Planet Symposium 2013, 9-13 Sep. 2013, Edinburgh, ESA Proceedings No. SP-722 (CD).

Abdelmoneim, H., Soliman, M. R., and Moghazy, H. M., 2020. Evaluation of TRMM 3B42V7 and CHIRPS Satellite Precipitation Products as an Input for Hydrological Model over Eastern Nile Basin. Earth Systems and Environment, 4(4), 685–698, doi: 10.1007/s41748-020-00185-3

Akinsanola, A. A., and Zhou, W., 2018. Projections of West African summer monsoon rainfall extremes from two CORDEX models. Climate Dynamics, 52(3), 2017-2028, DOI:10.1007/s00382-018-4238-8.

Akinsanola, A. A., Ogunjobi, K. O., Gbode, I. E., and Ajayi, V. O., 2015. Assessing the capabilities of three regional climate models over CORDEX Africa in simulating West African summer monsoon precipitation. Advances in Meteorology, doi:10.1155/2015/935431.

Akinsanola, A.A., Ajayi, V.O., Adejare, A.T., Adeyeri, O.E., Gbode, I.E., Ogunjobi, K.O., Nikulin, G. and Abolude, A.T., 2017. Evaluation of rainfall simulations over West Africa in dynamically downscaled CMIP5 global circulation models. Theoretical and applied climatology., 132(1), 437-450, doi:10.1007/s00704-017-2087-8.

Alexander, S., Atsebeha, E., Negatu, S., Kirksey, K., Brossard, D., Holzer, E., and Block, P., 2020. Development of an interdisciplinary, multi-method approach to seasonal climate forecast communication at the local scale, Climatic Change, 162, 2021–2042, DOI: 10.1007/s10584-020-02845-9.

Alexander, S., Yang, G., Addisu, G., and Block, P., 2021. Forecast-informed reservoir operations to guide hydropower and agriculture allocations in the Blue Nile basin, Ethiopia. International Journal of Water Resources Development, 1-26.

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., and Pappenberger, F., 2013. GloFAS–global ensemble streamflow forecasting and flood early warning. Hydrology and Earth System Sciences, 17(3), 1161-1175, doi:10.5194/hess-17-1161-2013.

Animashaun, I. M., Oguntunde, P. G., Akinwumiju, A. S., and Olubanjo, O. O., 2020. Rainfall Analysis over the Niger Central Hydrological Area, Nigeria: Variability, Trend, and Change point detection. Scientific African, 8, e00419, doi: 10.1016/j.sciaf.2020.e00419.

Arakawa, A., and Schubert, W. H., 1974. Interaction of a cumulus cloud ensemble with the large-scale environment, Part I. Journal of the Atmospheric Sciences, 31(3), 674-701.

Bartholmes, J., & Todini, E., 2005. Coupling meteorological and hydrological models for flood forecasting. Hydrol. Earth Syst. Sci. Discuss., 9, 333–346, doi:10.5194/hess-9-333-2005.

Barros, R., 2009. Integrated Water Resources Management in Mozambique: The case of the Limpopo Basin. Thesis submitted to the Faculty of Environmental Sciences. Federal Institute of Technology Zurich (ETH), Zurich, Switzerland. 85 pp.

Becker, E., Kirtman, B.P. and Pegion, K., 2020. Evolution of the North American multi-model ensemble. Geophysical Research Letters, 47(9), p.e2020GL087408, doi: 10.1029/2020GL087408

Belete, M., Deng, J., Wang, K., Zhou, M., Zhu, E., Shiferaw, E., and Bayissa, Y., 2020. Evaluation of satellite rainfall products for modeling water yield over the source region of Blue Nile Basin. Science of the Total Environment, 708, 134834. doi: 10.1016/j.scitotenv.2019.134834.

Benade, C., 1988. Episodic flood events in the Orange River system–an ecological perspective. In Paper 3.6 in Proceedings of the conference on floods in perspective (pp. 1-16). Pretoria: Department of Water Affairs and Forestry

Benjamin, S. G., Weygandt, S. S., Brown, J. M., Hu, M., Alexander, C. R., Smirnova, T. G., & Lin, H., 2016. A North American hourly assimilation and model forecast cycle: The rapid refresh. Mon. Weather Rev., 144, 1669–1694, doi:10.1175/MWR-D-15-0242.1.

Berner, J., Fossell, K. R., Ha, S. Y., Hacker, J. P., & Snyder, C., 2015. Increasing the skill of probabilistic forecasts: Understanding performance improvements from model-error representations. Mon. Weather Rev., 143, 1295–1320, doi: 10.1175/MWR-D-14-00091.1

Bhuiyan, M.A., Nikolopoulos. E.I., and Anagnostou, E.N., 2019. Machine Learning-Based Blending of Satellite and Reanalysis Precipitation Datasets: A Multiregional Tropical Complex Terrain Evaluation, Journal of Hydrometeorology, 20(11), 2147-2161, doi: 10.1175/JHM-D-19-0073.1.

Blake, E. S., & Zelinsky, D. A., 2017. (National Hurricane Center Tropical Cyclone Report - Hurricane Harvey (Pp. 1-43)).

Blaylock, B., Horel, J., & Liston, S., 2017. Cloud archiving and data mining of high resolution rapid refresh model output. Computers and Geosciences, 109, 43–50, doi: 10.1016/j.cageo.2017.08.005.

Bliefernicht, J., Waongo, M., Salack, S., Seidel, J., Laux, P., and Kunstmann, H., 2019. Quality and value of seasonal precipitation forecasts issued by the West African regional climate outlook forum. Journal of Applied Meteorology and Climatology, 58(3), 621-642, doi:10.1175/JAMC-D-18-0066.1.

Bougeault, P., Toth, Z., Bishop, C., Brown, B., Burridge, D., De Chen, H., … Worley, S., 2010. The thorpex interactive grand global ensemble. Bulletin of the American Meteorological Society, 91(8), 1059–1072, doi: 10.1175/2010BAMS2853.1

Breuer, N. E., Fraisse, C.W., and Cabrera, V.E., 2010. The Cooperative Extension Service as a boundary organization for diffu- sion of climate forecasts: A 5-year study. Journal of Extension 48(4).

Buizza, R., 2008. The value of probabilistic prediction. Atmospheric Science Letters, 9(2), 36–42, doi: 10.1002/asl.170

Bytheway, J. L., & Kummerow, C. D., 2015. Toward an object-based assessment of high-resolution forecasts of long-lived convective precipitation in the central US. Journal of Advances in Modeling Earth Systems, 7, 1248–1264, doi: 10.1002/2015MS000497.

Bytheway, J. L., Kummerow, C. D., & Alexander, C., 2017. A features-based assessment of the evolution of warm season precipitation forecasts from the HRRR model over three years of development. Weather and Forecasting, 32, 1841–1856, doi:10.1175/WAF-D-17-0050.1.

Cai, H., & Dumais, R. E., 2015. Object-based evaluation of a numerical weather prediction model's performance through forecast storm characteristic analysis. Weather and Forecasting, 30, 1451–1468, doi: 10.1175/WAF-D-15-0008.1.

Cambray, J.A., Davies, B.R., Ashton, P.J., Agnew, J.D., De Moor, F.C. and Skelton, P.H., 1986. The Orange-Vaal River system. In The ecology of river systems (pp. 89-161). Springer, Dordrecht, doi: 10.1007/978-94-017-3290-1_4

Ceppi, A., Ravazzani, G., Corbari, C., Salerno, R., Meucci, S., & Mancini, M., 2014. Real-time drought forecasting system for irrigation management. Hydrol. Earth Syst. Sci., 18, 3353–3366, doi: 10.5194/hess-18-3353-2014.

Cerrai, D., Yang, Q., Shen, X., Koukoula, M., & Anagnostou, E. N., 2020. Brief communication: Hurricane Dorian: Automated near-real-time mapping of the 'unprecedented' flooding on the Bahamas using SAR. Nat. Hazards Earth Syst. Sci., 20, 1463–1468., doi: 10.5194/nhess-2019-312

Chen, F., Janjić, Z., and Mitchell, K., 1997. Impact of atmospheric surface-layer parameterizations in the new land-surface scheme of the NCEP mesoscale Eta model. Boundary-Layer Meteorology, 85(3), 391-421.

Choong, S. M., & El-Shafie, A., 2015. State-of-the-art for modelling reservoir inflows and management optimization. Water resources management, 29(4), 1267-1282, doi: 10.1007/s11269-014-0872-z.

Chou, M.-D., & Suarez, M. J., 1994. An efficient thermal infrared radiation parameterization for use in general circulation models. NASA Tech. Memo., 85.

CGIAR. (2014, October 19). Limpopo River Basin Research Highlights. CGIAR Challenge Program on Water and Food., from https://waterandfood.cgiar.org/river-basins/limpopo-2/., Retrieved November 14, 2021

Clough, S., Shephard, M., Mlawer, E., Delamere, J., Iacono, M., Cady-Pereira, K., Boukabara, S., and Brown, P. 2005. Atmospheric radiative transfer MODELING: A summary of the AER codes. Journal of Quantitative Spectroscopy and Radiative Transfer, 91(2), 233-244. doi:10.1016/j.jqsrt.2004.05.058.

Collischonn, W., Tucci, C. E. M., Clarke, R. T., Chou, S. C., Guilhon, L. G., Cataldi, M., & Allasia, D., 2007. Medium-range reservoir inflow predictions based on quantitative precipitation forecasts. J. Hydrol., 344, 112–122, doi: 10.1016/j.jhydrol.2007.06.025.

Crimmins, M.A. and McClaran, M.P., 2016. Where do seasonal climate predictions belong in the drought management toolbox?. Rangelands, 38(4), pp.169-176, doi: 10.1016/j.rala.2016.06.004.

Cuo, L., Pagano, T. C., & Wang, Q. J., 2011. A review of quantitative precipitation forecasts and their use in short-to medium-range streamflow forecasting. Journal of Hydrometeorology, 12, 713–728, doi: 10.1175/2011JHM1347.1

De Wit, M., and Stankiewicz, J., 2006. Changes in surface water supply across Africa with predicted climate change. Science, 311 (5769), 1917-1921, doi: 10.1126/science.1119929.

Dembélé, M., Schaefli, B., van de Giesen, N., and Mariéthoz, G., 2020 Suitability of 17 rainfall and temperature gridded datasets for largescale hydrological modelling in West Africa. Hydrol. Earth Syst. Sci. Discuss., https://doi. org/10.5194/hess-2020-68, in review

Dezfuli, A. K., Ichoku, C. M., Huffman, G. J., Mohr, K. I., Selker, J. S., Van De Giesen, N., Hochreutener, R., and Annor, F. O., 2017a. Validation of IMERG precipitation in Africa. Journal of hydrometeorology, 18(10), 2817-2825, doi: 10.1175/JHM-D-17-0139.1.

Dezfuli, A. K., Ichoku, C. M., Mohr, K. I., & Huffman, G. J., 2017b Precipitation characteristics in West and East Africa from satellite and in situ observations. Journal of Hydrometeorology, 18(6), 1799-1805, doi:10.1175/JHM-D-17-0068.1.

Di Baldassarre, G., Elshamy, M., van Griensven, A., Soliman, E., Kigobe, M., Ndomba, P., Mutemi, J., Mutua, F., Moges, S., Xuan, Y., Solomatine, D., and Uhlenbrook, S., 2011.

Future hydrology and climate in the River Nile basin: a review. Hydrological Sciences Journal, 56:2, 199-211, DOI: 10.1080/02626667.2011.557378.f

Djaman, K., Balde, A.B., Rudnick, D.R., Ndiaye, O., and Irmak, S., 2017. Long-term trend analysis in climate variables and agricultural adaptation strategies to climate change in the Senegal River Basin. Int. J. Clim., 37(6), 2873-2888, doi: 10.1002/joc.4885.

Dutra, E., Magnusson, L., Wetterhall, F., Cloke, H.L., Balsamo, G., Boussetta, S., and Pappenberger, 2013. The 2010-2011 drought in the Horn of Africa in ECMWF reanalysis and seasonal forecast products, International Journal of Climatology, 33, 1720-1729, DOI: 10.1002/joc.3545.

Ebert, E. E., Damrath, U., Wergen, W., & Baldwin, M. E., 2003. Supplement to The WGNE Assessment of Short-term Quantitative Precipitation Forecasts. Bulletin of the American Meteorological Society, 84(4), 492–492, doi: 10.1175/BAMS-84-4-481.

Ebert, E. E., Janowiak, J. E., & Kidd, C., 2007. Comparison of Near-Real-Time Precipitation Estimates from Satellite Observations and Numerical Models. Bulletin of the American Meteorological Society, 88(1), doi: 10.1175/BAMS-88-1-47.

Ebert, E. E., & McBride, J. L., 2000. Verification of precipitation in weather systems: Determination of systematic errors. Journal of Hydrology, 239(1–4), 179–202, doi: 10.1016/S0022-1694(00)00343-7

FAO, 2004. Drought Impact Mitigation and Prevention in the Limpopo River Basin: A Situation Analysis; Land and Water Discussion Paper No. 40; Food and Agriculture Organization of the United Nations: Roma, Italy, 2004, Fenta, A. A., Yasuda, H., Shimizu, K., Ibaraki, Y., Haregeweyn, N., Kawai, T., Belay, A.S., Sultan, D. and Ebabu, K., 2018. Evaluation of satellite rainfall estimates over the Lake Tana basin at the source region of the Blue Nile River. Atmospheric Research, 212, 43-53.

Feaster, T. D., Weaver, J. C., Gotvald, A. J., & Kolb, K. R., 2018. Preliminary Peak Stage and Streamflow Data for Selected US Geological Survey Streamgaging Stations in North and South Carolina for Flooding Following Hurricane Florence, September 2018 (No. 2018-1172). US Geological Survey.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A. and Michaelsen, J., 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Scientific data, 2(1), 1-21.

Gaye, C. B., Diaw, M., & Malou, R., 2013. Assessing the impacts of climate change on water resources of a West African trans-boundary river basin and its environmental consequences (Senegal River Basin). Sci. Cold Arid Reg, 5(1), 0140-0156, doi: 10.3724/SP.J.1226.2013.00140.

Gebregiorgis, A. S., Kirstetter, P. E., Hong, Y. E., Carr, N. J., Gourley, J. J., Petersen, W., & Zheng, Y., 2017. Understanding overland multisensor satellite precipitation error in TMPA-RT products. Journal of Hydrometeorology, 18, 285–306, doi: 10.1175/JHM-D-15-0207.1.

Gebremichael, N., Bitew, M.M., Firpa, H.A., and Tesfay, G., 2014. Accuracy of satellite rainfall estimates in the Blue Nile Basin: Lowland plain versus highland mountain. Water Resources Research, 50(11), 8775-8790, doi: 10.1002/2013WR014500.

Golding, B. W., 2000. Quantitative precipitation forecasting in the UK. Journal of Hydrology, 239(1–4), 286–305, doi: 10.1016/S0022-1694(00)00354-1

Gossett, M., Alcoba, M., Roca, R., Cloche, S., and Urbani, G., 2018. Evaluation of TAPEER daily estimates and other GPM-era products against dense gauge networks in West Africa, analysing ground reference uncertainty. 144(1), 255-269, doi: 10.1002/qj.3335.

Gourley, J. J. & others., 2017. The FLASH Project: Improving the tools for flash flood monitoring and prediction across the United States. Bulletin of the American Meteorological Society, 98, 361–372, doi: 10.1175/bams-d-15-00247.1

Gowan, T. M., Steenburgh, W. J., & Schwartz, C. S., 2018. Validation of mountain precipitation forecasts from the convection-permitting NCAR ensemble and operational forecast systems over the Western United States. Weather and Forecasting, 33, 739–765, doi: 10.1175/WAF-D-17-0144.1

Grell, G. A., 1993. Prognostic evaluation of assumptions used by cumulus parameterizations. Monthly weather review, 121(3), 764-787.

Grosenick, G., Djegal, A. and King, J.W., 1990. Analyse pour la gestion des ressources naturelles du Sénégal: rapport final. Institute for Development Anthropology, Washington, DC

Gubler, S., Sedlmeier, K., Bhend, J., Avalos, G., Coelho, C. A. S., Escajadillo, Y., Y., Jacques-Coper, M., Martinez, R., Schwierz, C., De Skansi, M., and Spirig, C., 2020. Assessment of ECMWF SEAS5 seasonal forecast performance over South America. Weather and Forecasting, 35(2), 561–584, doi: 10.1175/WAF-D-19-0106.1.

Gumindoga, W., Rientjes, T. H., Haile, A. T., Makurira, H., & Reggiani, P., 2019. Performance of bias-correction schemes for CMORPH rainfall estimates in the Zambezi River basin. Hydrology and earth system sciences, 23(7), 2915-2938, doi: 10.5194/hess-23-2915-2019.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of hydrology, 377(1-2), 80-91, doi: 10.1016/j.jhydrol.2009.08.003.

Haiden, T., Janousek, M., Vitart, F., Ben-Bouallegue, Z., Ferranti, L., Prates, C., and Richardson, D. 2021. Evaluation of ECMWF forecasts, including the 2020 upgrade. ECMWF Technical, 80, doi: 10.21957/6njp8byz4

Haile, A. T., Tefera, F. T., and Rientjes, T., 2016. Flood forecasting in Niger-Benue basin using satellite and quantitative precipitation forecast data. International journal of applied earth observation and geoinformation, 52, 475-484, doi: 10.1016/j.jag.2016.06.021.

He, H., Yao, S., Huang, A., & Gong, K. (2020). Evaluation and Error Correction of the ECMWF Subseasonal Precipitation Forecast over Eastern China during Summer. Advances in Meteorology, 2020, doi:10.1155/2020/1920841

Heimann, D. C., Holmes, R. R. Jr., and Harris, T. E., 2018. Flooding in the Southern Midwestern United States, April–May 2017 (No. 2018-1004)

Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., Oki, R., Nakamura, K., and Iguchi, T., 2014. The global precipitation measurement mission. Bulletin of the American Meteorological Society, 95(5), 701-722.

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Xie, P., & Yoo, S. H., 2015. NASA global precipitation measurement (GPM) integrated multi-satellite retrievals for GPM (IMERG). Algorithm Theoretical Basis Document (ATBD) Version, 4, 26.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K.P. and Stocker, E. F., 2007. The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation estimates at fine scales. Journal of Hydrometeorology, 8(1), 38-55. doi:10.1175/jhm560.1

Huffman, G.J., Stocker, E.F., Bolvin, D.T., Nelkin, E.J., Jackson J., 2019a. GPM IMERG Final Precipitation L3 1 day 0.1 degree x 0.1 degree V06, Edited by Andrey Savtchenko, Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: [Feb 11, 2021], 10.5067/GPM/IMERGDF/DAY/06

Huffman, G.J., Stocker, E.F., Bolvin, D.T., Nelkin, E.J., Jackson J., 2019b. GPM IMERG Early Precipitation L3 1 day 0.1 degree x 0.1 degree V06, Edited by Andrey Savtchenko, Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: [Feb 11, 2021], 10.5067/GPM/IMERGDE/DAY/06

Iacono, M. J., Mlawer, E. J., Clough, S. A., and Morcrette, J. J., 2000. Impact of an improved longwave radiation model, RRTM, on the energy budget and thermodynamic properties of the NCAR community climate model, CCM3. Journal of Geophysical Research: Atmospheres, 105(D11), 14873-14890, doi: 10.1029/2000JD900091.

Ikeda, K., Steiner, M., Pinto, J., & Alexander, C., 2013. Evaluation of cold-season precipitation forecasts generated by the hourly updating high-resolution rapid refresh model. Weather and Forecasting, 28, 921–939, doi: 10.1175/WAF-D-12-00085.1

Ikeda, K., Steiner, M., & Thompson, G., 2017. Examination of mixed-phase precipitation forecasts from the high-resolution rapid refresh model using surface observations and sounding data. Weather and Forecasting, 32, 949–967, doi: 10.1175/WAF-D-16-0171.1.

Jehanzaib, M., Bilal Idrees, M., Kim, D., & Kim, T. W., 2021. Comprehensive Evaluation of Machine Learning Techniques for Hydrological Drought Forecasting. Journal of Irrigation and Drainage Engineering, 147(7), 04021022, doi: 10.1061/(ASCE)IR.1943-4774.0001575.

Jiang, M., Feng, J., Li, Z., Sun, R., Hou, Y. T., Zhu, Y., Wan, B., Guo, J., and Cribb, M., 2017. Potential influences of neglecting aerosol effects on the NCEP GFS precipitation forecast. Atmospheric Chemistry and Physics, 17(22), 13967–13982, doi: 10.5194/acp-17-13967-2017.

JMA, 2019. Outline of the operational numerical weather prediction at the Japan Meteorological Agency (Appendix to WMO numerical weather prediction progress report). Japan Meteorological Agency, 47pp. (available online at https://www.jma.go.jp/jma/jma-eng/jma-center/nwp/outline2019-nwp/index.htm; last accessed: February 2021)

Kim, D., Nelson, B., & Seo, D.-J., 2009. Characteristics of Reprocessed Hydrometeorological Automated Data System (HADS) Hourly Precipitation Data. Weather and Forecasting, 24, 1287–1296, doi: 10.1175/2009waf2222227.1

Kling, H., Fuchs, M., and Paulin, M., 2012. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. Journal of Hydrology, 424, 264-277. Doi: 10.1016/j.jhydrol.2012.01.011

Kling, H., Stanzel, P., & Preishuber, M., 2014. Impact modelling of water resources development and climate scenarios on Zambezi River discharge. Journal of Hydrology: Regional Studies, 1, 17-43, doi: 10.1016/j.ejrh.2014.05.002.

Koppa, A., Gebremichael, M., Zambon, R.C., Yeh, W.W.G., and Hopson, T.M., 2019. Seasonal Hydropower Planning for Data-Scarce Regions Using Multimodel Ensemble Forecasts, Remote Sensing Data, and Stochastic Programming. WATER RESOURCES RESEARCH, 55(11), 8583-8607, DOI: 10.1029/2019WR025228.

Lange, G.M., Mungatana, E., and Hassan, R., 2007. Water accounting for the Orange River Basin: An economic perspective on managing a transboundary resource. Ecological Economics, 61(4), 660-670, doi: 10.1016/j.ecolecon.2006.07.032.

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J. C., Rödel, R., Sindorf, N., & Wisser, D., 2011. High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. Frontiers in Ecology and the Environment, 9(9), 494–502, doi: 10.1890/100125.

Lehner, B., Verdin, K., Jarvis, A., 2008. New global hydrography derived from spaceborne elevation data. Eos, Transactions, AGU, 89(10): 93-94.

Lien, G. Y., Kalnay, E., Miyoshi, T., and Huffman, G. J., 2016. Statistical properties of global precipitation in the NCEP GFS model and TMPA observations for data assimilation. Monthly Weather Review, 144(2), 663–679, doi: 10.1175/MWR-D-15-0150.1.

Lin, H., Merryfield, W. J., Muncaster, R., Smith, G., Markovic, M., Erfani, A., et al., (2019). The Canadian seasonal to interannual prediction system version 2. (CanSIPSv2). https://iridl.ldeo.columbia.edu/documentation/Models/NMME/CanSIPSv2/technote_cansips-v2_ (2019)0703_e.pdf.

Maranan, M., Fink, A. H., Knippertz, P., Amekudzi, L. K., Atiah, W. A., & Stengel, M., 2020. A Process-Based Validation of GPM IMERG and Its Sources Using a Mesoscale Rain Gauge Network in the West African Forest Zone. Journal of Hydrometeorology, 21(4), 729–749, doi: 10.1175/JHM-D-19-0257.1

Mase, A. S., and Prokopy, L.S., 2014. Unrealized Potential: A review of perceptions and use of weather and climate infor- mation in agricultural decision making. Weather Climate and Society 6(1):47-61. DOI: 10.1175/Wcas-D-12-00062.1.

Merryfield, W.J., Lee, W.S., Boer, G.J., Kharin, V.V., Scinocca, J.F., Flato, G.M., Ajayamohan, R.S., Fyfe, J.C., Tang, Y. and Polavarapu, S., 2013. The Canadian seasonal to interannual prediction system. Part I: Models and initialization. Monthly weather review, 141(8), pp.2910-2945, doi: 10.1175/MWR-D-12-00216.1.

Mohamed, A. E., 2001. Management of International Water Resources: A Case Study of Legal and Institutional Aspects of the Limpopo and Orange River Basins in Southern Africa. Kungl. Tekniska högskolan. 33 pp.

Molod, A., Hackert, E., Vikhliaev, Y., Zhao, B., Barahona, D., Vernieres, G., et al., 2020. GEOS-S2S Version 2: The GMAO high resolution coupled model and assimilation system for seasonal prediction. Journal of Geophysical Research: Atmospheres, 125, e2019JD031767, doi:10.1029/2019JD031767

Molteni, F., Stockdale, T., Balmaseda, M., Balsamo, G., Buizza, R., Ferranti, L., Magnusson, L., Mogensen, K., Palmer, T. and Vitart, F., 2011. The new ECMWF seasonal forecast system (System 4) (Vol. 49). Reading: European Centre for medium-range weather forecasts.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., and Clough, S. A., 1997. Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. Journal of Geophysical Research: Atmospheres, 102(D14), 16663-16682.

Morss, R. E., 2003. Assessing the needs of users of warm season quantitative precipitation forecasts in Colorado. Symp. on Impacts of Water Variability: Benefits and Challenges. Preprints, 83rd AMS Annual Meeting, 9–13.

Nakanishi, M., & Niino, H., 2004. An Improved Mellor–Yamada Level-3 Model with Condensation Physics: Its Design and Verification. Boundary-Layer Meteorology, 112, 1–31, doi: 10.1023/b:boun.0000020164.04146.98.

National Centers for Environmental Prediction, 2015. NCEP GFS 0.25 Degree Global Forecast Grids Historical Archive, doi: 10.5065/D65D8PWK.

National Centers for Environmental Prediction, 2019a. National Centers for Environmental Prediction: The Global Forecast System (GFS) -Global Spectral Model (GSM). Retrieved from https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gfs/documentation.php, last access: Sep 2020.

National Centers for Environmental Prediction, 2019b. National Centers for Environmental Prediction: FV3: The GFDL Finite-Volume Cubed-Sphere Dynamical Core, Retrieved from https://vlab.ncep.noaa.gov/web/fv3gfs, last access: Sep 2020).

National Centers for Environmental Prediction, 2021a. The Global Forecast System (GFS) – Global Spectral Model, Retrieved from https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gfs/documentation_spectralgfs.php, last accessed: June 2021).

National Centers for Environmental Prediction. 2021b. The Global Forecast System (GFS) Documentation, Retrieved from https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gfs/documentation.php, last accessed: June 2021).

NOAA (National Oceanic and Atmospheric Administration)., 2016. National Water Model: Improving NOAA's Water Prediction Services, url: https://water.noaa.gov/documents/wrn-national-water-model.pdf

NBI, 2012a. Climate change and its implications for the Nile region. In: State of the River Nile Basin. Nile Basin Initiative, Uganda, pp 205–224. ISSN:978-9970-444-00-7.

NBI, 2012b. The water resources of the Nile basin. In: State of the River Nile Basin. Nile Basin Initiative, Uganda, pp 25–56, ISSN:978-9970-444-00-7.

Onyutha C., Tabari H., Taye M.T., Nyandwaro G.N., and Willems, P., 2016. Analyses of rainfall trends in the Nile river basin. J. Hydro. Environ. Res., 13, 36–51, DOI: 10.1016/j.jher.2015.09.002.

Pandya, R., Hodgson, A., Hayden, M.H., Akweongo, P., Hopson, T., Forgor, A.A., Yoksas, T., Dalaba, M.A., Dukic, V., Mera, R., and Dumont, A., McCormack, K., Anaseba, D., Awine, T., Boehnert, J., Nyaaba, G., Laing, A., and Semazzi, F., 2015. Using weather forecasts to help manage meningitis in the West African Sahel. Bulletin of the American Meteorological Society 96(1):103-+. DOI: 10.1175/Bams-D-13-00121.1.

Patt, A. G., Ogallo, L., and Hellmuth, M., 2007. Sustainability—Learning from 10 years of climate outlook forums in Africa. Science 318(5847):49-50. DOI: 10.1126/science.1147909.

Pinto, J. O., Grim, J. A., & Steiner, M., 2015. Assessment of the High-Resolution Rapid Refresh Model's Ability to Predict Mesoscale Convective Systems Using Object-Based Evaluation. Weather and Forecasting, 30, 892–913, doi: 10.1175/waf-d-14-00118.1.

Pirret, J. S., Daron, J. D., Bett, P. E., Fournier, N., and Foamouhoue, A. K., 2020. Assessing the skill and reliability of seasonal climate forecasts in Sahelian West Africa. Weather and Forecasting, 35(3), 1035-1050, doi:10.1175/WAF-D-19-0168.1.

Prasad, R., Deo, R. C., Li, Y., & Maraseni, T., 2019. Weekly soil moisture forecasting with multivariate sequential, ensemble empirical mode decomposition and Boruta-random forest hybridizer algorithm approach. Catena, 177, 149-166, doi: 10.1016/j.catena.2019.02.012.

Rauch, M., Bliefernicht, J., Laux, P., Salack, S., Waongo, M., & Kunstmann, H., 2019. Seasonal Forecasting of the Onset of the Rainy Season in West Africa. Atmosphere, 10(9), 528, doi: 10.3390/atmos10090528

Roudier, P., Alhassane, A., Baron, C., Louvet, S., and Sultan, B., 2016. Assessing the benefits of weather and seasonal forecasts to millet growers in Niger. Agricultural and forest meteorology, 223, 168-180, doi: 10.1016/j.agrformet.2016.04.010.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H.-Y., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M.P., van den Dool, H., Zhang, Q., Wang, W., Chen, M., and Becker, E., 2014. The NCEP climate forecast system version 2. Journal of climate, 27(6), 2185-2208.

Satgé, F., Defrance, D., Sultan, B., Bonnet, M. P., Seyler, F., Rouché, N., and Paturel, J. E., 2020. Evaluation of 23 gridded precipitation datasets across West Africa. Journal of Hydrology, 581, 124412, doi:10.1016/j.jhydrol.2019.124412 .

Sahlu, D., Nikolopoulos, E. I., Moges, S. A., Anagnostou, E. N., and Hailu, D., 2016. First evaluation of the Day-1 IMERG over the upper Blue Nile basin. Journal of Hydrometeorology, 17(11), 2875-2882.

Schneider, J. M., & Garbrecht, J. D., 2003. A measure of the usefulness of seasonal precipitation forecasts for agricultural applications. Transactions of the ASAE, 46(2), 257, doi: 10.13031/2013.12976.

Seo, B. C., Quintero, F., & Krajewski, W. F., 2018. High-resolution QPF uncertainty and its implications for flood prediction: A case study for the eastern Iowa flood of 2016. Journal of Hydrometeorology, 19, 1289–1304, doi: 10.1175/JHM-D-18-0046.1.

Smalley, M., Kirstetter, P. E., & L'Ecuyer, T., 2017. How frequent is precipitation over the contiguous United States? Perspectives from ground-based and spaceborne radars. Journal of Hydrometeorology, 18, 1657–1672, doi: 10.1175/JHM-D-16-0242.1

Smirnova, T. G., Brown, J. M., & Benjamin, S. G., 1997. Evolution of soil moisture and temperature in the MAPS/RUC assimilation cycle. 13th Conf. on Hydrology, 172–175.

Sorí, R., Nieto, R., Drumond, A., and Gimeno, L., 2017. The Niger River Basin Moisture Sources: A Lagrangian Analysis. Atmosphere, 8(2), 38, doi: 10.3390/atmos8020038.

Sridevi, C., Singh, K. K., Suneetha, P., Durai, V. R., and Kumar, A., 2018. Rainfall forecast skill of Global Forecasting System (GFS) model over India during summer monsoon 2015. Geofizika, 35(1), 39–52, doi: 10.15233/gfz.2018.35.4.

Stellingwerf, S., Riddle, E., Hopson, T., Knievel, Brown, B., and Gebremichael, M., 2021. Optimizing precipitation forecasts for hydrological catchments in Ethiopia using statistical bias correction and multi-modeling, Earth and Space Science, doi: 10.1029/2019EA000933.

Sukovich, E. M., Ralph, F. M., Barthold, F. E., Reynolds, D. W., & Novak, D. R., 2014. Extreme quantitative precipitation forecast performance at the weather prediction center from 2001 to 2011. Weather and Forecasting, 29, 894–911, doi: 10.1175/WAF-D-13-00061.1.

Sutcliffe, J. V., and Parks, Y. P., 1999. The hydrology of the Nile, IAHS Special Publication no. 5,. Wallingford, UK: International Association of Hydrological Sciences.

Swinbank, R., Kyouda, M., Buchanan, P., Froude, L., Hamill, T.M., Hewson, T.D., Keller, J.H., Matsueda, M., Methven, J., Pappenberger, F. and Scheuerer, M., 2016. The TIGGE project and its achievements. Bulletin of the American Meteorological Society, 97(1), pp.49-67, doi: 10.1175/BAMS-D-13-00191.1.

Sylla, M. B., Faye, A., Giorgi, F., Diedhiou, A., and Kunstmann, H., 2018. Projected heat stress under 1.5 C and 2 C global warming scenarios creates unprecedented discomfort for humans in West Africa. Earth's Future, 6(7), 1029-1044, doi: 10.1029/2018EF000873.

Tang, G., Zhou, H., Li, N., Wang, F., Wang, Y., & Jian, D., 2010. Value of Medium-range Precipitation Forecasts in Inflow Prediction and Hydropower Optimization. Water Resources Management, 24(11), 2721–2742, doi:10.1007/s11269-010-9576-1.

Tian, D., Wood, E.F., and Yuan, X., 2017. CFSv2-based sub-seasonal precipitation and temperature forecast skill over the contiguous United States. Hydrology and Earth System Sciences, 21, 1477–1490, doi:10.5194/hess-21-1477-2017.

Thiemig, V., Rojas, R., Zambrano-Bigiarini, M., Levizzani, V., & De Roo, A., 2012. Validation of satellite-based precipitation products over sparsely gauged African river basins. Journal of Hydrometeorology, 13(6), 1760-1783, doi: 10.1175/JHM-D-12-032.1

Thielen, J., Bartholmes, J., Ramos, M.-H., & de Roo, A., 2009. The European Flood Alert System – Part 1: Concept and development. Hydrology and Earth System Sciences, 13(2), 125–140, doi: 10.5194/hess-13-125-2009

Thompson, G., Field, P. R., Rasmussen, R. M., & Hall, W. D., 2008. Explicit Forecasts of Winter Precipitation Using an Improved Bulk Microphysics Scheme. Part II: Implementation of a New Snow Parameterization. Monthly Weather Review, 136, 5095–5115, doi: 10.1175/2008mwr2387.1.

Towner, J., Cloke, H.I., Zsoter, E., Flamig, Z., Hoch, J.M., Bazo, J., de Perez, E.C., and Stephens, E.M., 2019. Assessing the performance of global hydrological models for capturing peak river flows in the Amazon basin. Hydrology and Earth System Sciences, 23, 3057–3080, doi:10.5194/hess-23-3057-2019.

Trenberth, K.E. and Trenberth, K.E. eds., 1992. Climate system modeling. Cambridge University Press.

van de Giesen, N., Hut, R., & Selker, J., 2014. The trans-African hydro-meteorological observatory (TAHMO). Wiley Interdisciplinary Reviews: Water, 1(4), 341-348, doi:10.1002/wat2.1034.

Valeriano, O. C. S., Koike, T., Yang, K., Graf, T., Li, X., Wang, L., & Han, X., 2010. Decision support for dam release during floods using a distributed biosphere hydrological model driven by quantitative precipitation forecasts. Water Resour. Res., 46, W10544, doi: 10.1029/2010WR009502

Wang, J. W. A., Sardeshmukh, P. D., Compo, G. P., Whitaker, J. S., Slivinski, L. C., McColl, C. M., and Pegion, P. J., 2019. Sensitivities of the NCEP global forecast system. Monthly Weather Review, 147(4), 1237–1256, doi: 10.1175/MWR-D-18-0239.1.

Wang, T., Chu, C., Sun, X., & Li, T., 2020. Improving real-time forecast of intraseasonal variabilities of Indian summer monsoon precipitation in an empirical scheme. Frontiers in Earth Science, 408 doi: 10.3389/feart.2020.577311.

Watson, K. M., Harwell, G. R., Wallace, D. S., Welborn, T. L., Stengel, V. G., & McDowell, J. S., 2018. Characterization of Peak Streamflows and Flood Inundation of Selected Areas in Southeastern Texas and Southwestern Louisiana from the August and September 2017 Flood Resulting from Hurricane Harvey (No. 2018-5070). US Geological Survey.

Wernli, H., Paulat, M., Hagen, M., & Frei, C., 2008. SAL—a novel quality measure for the verification of quantitative precipitation forecasts. Mon. Weather Rev., 136, 4470–4487, doi.org/10.1175/2008MWR2415.1.

Xu, L., Chen, N., Zhang, X., Chen, Z., Hu, C., & Wang, C., 2019. Improving the North American multi-model ensemble (NMME) precipitation forecasts at local areas using wavelet and machine learning. Climate dynamics, 53(1), 601-615, doi: 10.1007/s00382-018-04605-z.

Yu, P. S., Yang, T. C., Chen, S. Y., Kuo, C. M., & Tseng, H. W., 2017. Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting. Journal of Hydrology, 552, 92-104, doi: 10.1016/j.jhydrol.2017.06.020.

Yuan, X., E. F. Wood, and M. Liang., 2014. Integrating weather and climate prediction: toward seamless hydrologic forecasting. Geophysical Research Letters, 41, 5891–5896, doi:10.1002/2014GL061076.

Yue, H. and Gebremichael, M., 2020. Evaluation of high-resolution rapid refresh (HRRR) forecasts for extreme precipitation. Environmental Research Communications, 2(6), p.065004. DOI: 10.1088/2515-7620/ab9002.

Yue, H., Gebremichael, M. and Nourani, V., 2021a. Evaluation of Global Forecast System (GFS) Medium-Range Precipitation Forecasts in the Nile River Basin. Journal of Hydrometeorology. accepted. doi: 10.1175/JHM-D-21-0110.1.

Yue, H., Gebremichael, M. and Nourani, V., 2021b. Performance of the Global Forecast System's Medium-Range Precipitation Forecasts in the Niger River Basin. Hydrology and Earth System Sciences Discussions, pp.1-31. Accepted, doi: 10.5194/hess-2021-250.

Zhang, C., Xue, M., Supinie, T. A., Kong, F., Snook, N., Thomas, K. W., Brewster, K., Jung, Y., Harris, L.M. and Lin, S. J., 2019. How well does an FV3-based model predict precipitation at a convection-allowing resolution? Results from CAPS forecasts for the 2018 NOAA hazardous weather test bed with different physics combinations. Geophysical Research Letters, 46(6), 3523-3531, doi:10.1029/2018GL081702.

Zhang, J., Howard, K., Langston, C., Kaney, B., Qi, Y., Tang, L., & Arthur, A., 2016. Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. Bull. Am. Meteorol. Soc., 97, 621–638, doi: 10.1175/BAMS-D-14-00174.1.

Zhang, J., Howard, K., Langston, C., Vasiloff, S., Kaney, B., Arthur, A., & Seo, D. J., 2011. National mosaic and multi-sensor QPE (NMQ) system: Description, results, and future plans. Bull. Am. Meteorol. Soc., 92, 1321–1338, doi: 10.1175/2011BAMS-D-11-00047.1.