Predictability of rainfall over Equatorial East Africa from daily to sub-monthly time scales using multi-model ensemble forecasts

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

(Dr. rer. nat.)

von der KIT-Fakultät für Physik des

Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

M.Sc. Simon Ageet

aus Ngora

Tag der mündlichen Prüfung:	28.07.2023
Referent:	Prof. Dr. Andreas H. Fink
Korreferent:	Prof. Dr. Peter Knippertz

Abstract

The people of Equatorial East Africa (EEA) are heavily reliant on rain-fed agriculture for their livelihoods. For some of the countries, more than 35% of the GDP comes from the agricultural sector. Moreover, this region's communities are some of the most vulnerable to impacts of extreme weather, such that accurate weather information is crucial for the region. However, due to low skill, uptake of weather forecast information is still low. The aim of this thesis is to improve the understanding of the quality of forecasts over EEA by verifying them with the best available observations, with the hope of improving the uptake and use.

To start with, the quality of satellite rainfall estimates (SREs) was evaluated to select the most suitable product. SREs were the best option to use as the observation data set given the sparsity of the rain gauge (RG) network over EEA. Five gauge-calibrated SREs, namely, Integrated Multi-satellitE Retrieval for Global Precipitation Measurement (GPM; IMERG), Tropical Rainfall Measuring Mission (TRMM) Multisatellite precipitation Analysis (TMPA), Climate Hazard Infrared Precipitation with Stations (CHIRPS), and Multi-Source Weighted-Ensemble Precipitation (MSWEP; v2.2 and v2.8), were evaluated using a unique daily rainfall dataset from 36 stations across EEA for the period 2001–18. Based on the results, the SREs reproduce the annual rainfall pattern and seasonal rainfall cycle well, despite exhibiting biases in rainfall amount of up to 9%. For daily, pentadal and decadal rainfall accumulations, IMERG is the best product, while MSWEPv2.2 and CHIRPS perform best at the monthly and annual time steps, respectively. Rather disappointingly, all the SREs miss over 79% of daily extreme rainfall events recorded by the rain gauges. From this analysis, IMERG was chosen as the satellite observation going forward.

In the next step, ensemble reforecasts from European Centre for Medium-Range Weather Forecasts (ECMWF) were validated and postprocessed using IMERG and RG data for the period 2001–2018. The reforecasts were analysed from short to medium-range time scales, and for a range of temporal aggregations, i.e., 24, 48, and 120 hours. The skill was assessed using a reference forecast in the form of an extended probabilistic climatology (EPC) derived from the observations. Results show that the reforecasts overestimate rainfall, especially during the rainy seasons and over high altitude areas. However, there is potential skill in the raw forecasts up to 14 days ahead. There is

also an improvement of up to 30% in Brier score/continuous rank probability score relative to EPC in most areas, especially the higher-altitude regions, decreasing with lead-time. Aggregating the reforecasts enhances the skill further, likely due to reduction in timing mismatches. However, for some regions of the study domain, the predictive performance is worse than that of EPC, mainly due to biases in amount and dispersion errors. Postprocessing the reforecasts using isotonic distributional regression (IDR) considerably improves skill, increasing the number of grid-points with positive Brier skill score (continuous rank probability score) by 82% (67%) at 1-day lead-time.

Finally, regime-based analyses were carried out to assess whether weather regimes influence the rainfall and skill of the reforecasts relative to the reference forecast, EPC. Specifically, the Madden-Julian Oscillation (MJO), Kelvin wave and the Indian Ocean Dipole (IOD) were analysed. The results show that MJO and Kelvin waves strongly modulate rainfall by ~5–7 mmd^{-1}. The modulation is positive (negative) during the wet (dry) phases. The modulation is mostly over Uganda and the East African Coast for MJO and the Congo basin for Kelvin waves. The skill of the reforecasts is improved during the wet phases for rainfall detection and accuracy of rainfall amounts. The IOD also positively modulates rainfall and skill of the reforecasts during the positive, i.e. wet, phase while doing the opposite during the dry phase. The work in this thesis has highlighted the level of skill of ECMWF reforecasts, the benefit of postprocessing forecasts, and the potential sources of predictability, serving as a basis for forecast quality improvement over the region.

Kurzfassung

Die Menschen in Äquatorial-Ostafrika (EWR) sind für ihren Lebensunterhalt in hohem Maße von der Regenfeldwirtschaft abhängig. In einigen Ländern stammen mehr als 35% des BIP aus dem Agrarsektor. Darüber hinaus gehören die Menschen in dieser Region zu denjenigen, die am stärksten von den Auswirkungen extremer Wetterereignisse betroffen sind, so dass genaue Wetterinformationen für die Region von entscheidender Bedeutung sind. Die Nutzung von Wettervorhersagen ist jedoch aufgrund der geringen Fachkenntnisse nach wie vor gering. Ziel dieser Arbeit ist es, das Verständnis für die Qualität der Vorhersagen in EWR zu verbessern, indem die Vorhersagen mit den besten verfügbaren Beobachtungen überprüft werden, in der Hoffnung, die Akzeptanz und Nutzung zu verbessern.

Zunächst wurde die Qualität der satellitengestützten Niederschlagsschätzungen (SREs) bewertet, um das am besten geeignete Produkt auszuwählen. Aufgrund des spärlichen Netzes von Regenmessern (RG) im EWR waren SREs die beste Option für die Verwendung als Beobachtungsdatensatz. Fünf mit Messgeräten kalibrierte SREs, IMERG, TMPA, CHIRPS und MSWEP (v2.2 und v2.8), wurden anhand eines einzigartigen täglichen Niederschlagsdatensatzes von 36 Stationen im gesamten EWR für den Zeitraum 2001-18 bewertet. Die Ergebnisse zeigen, dass die SREs das jährliche Niederschlagsmuster und den saisonalen Niederschlagszyklus gut wiedergeben, obwohl sie Abweichungen von bis zu 9% bei der Niederschlagsmenge aufweisen. Für die tägliche, pentadische und dekadische Niederschlagsakkumulation ist IMERG das beste Produkt, während MSWEPv2.2 und CHIRPS bei den monatlichen bzw. jährlichen Zeitschritten am besten abschneiden. Enttäuschend ist, dass alle SREs über 79% der von den Regenmessern aufgezeichneten täglichen extremen Niederschlagsereignisse nicht erfassen. Aufgrund dieser Analyse wurde IMERG als künftige Satellitenbeobachtung ausgewählt.

In einem nächsten Schritt wurden die Ensemble-Vorhersagen des ECMWF für den Zeitraum 2001-2018 validiert und mit IMERG- und RG-Daten nachbearbeitet. Die Vorhersagen wurden auf kurz- bis mittelfristigen Zeitskalen und für eine Reihe von zeitlichen Aggregationen (stündlich, 48 Stunden und 120 Stunden) analysiert. Der Skill wurde anhand einer Referenzvorhersage in Form einer erweiterten probabilistischen Klimatologie (EPC) bewertet, die aus den Beobachtungen abgeleitet wurde. Die Ergebnisse zeigen, dass die Vorhersagen die Niedershlagsmenge überschätzen, insbesondere während

der Regenzeit und über hoch gelegenen Gebieten. Trotzdem können die Prognosen bis zu 14 Tage im Voraus verbessert werden. In den meisten Gebieten, vor allem in den höher gelegenen Regionen, ergibt sich eine Verbesserung von bis zu 30% der Brier scores bzw. der Werte für die kontinuierliche Rangwahrscheinlichkeit im Vergleich zu den EPC-Werten, die mit der Vorlaufzeit abnimmt. Die Aggregation der Vorhersagen verbessert die Fähigkeit weiter, was wahrscheinlich auf die Verringerung der zeitlichen Diskrepanzen zurückzuführen ist. Für einige Regionen des Untersuchungsgebiets ist die Vorhersagequalitat jedoch schlechter als die des EPC, was hauptsächlich auf Verzerrungen bei den Mengen- und Ausbreitungsfehlern zurückzuführen ist. Die Nachbearbeitung der Vorhersagen mittels isotonischer Verteilungsregression (IDR) verbessert die Qualität beträchtlich und erhöht die Anzahl der Gitterpunkte mit positivem Brier-Skill-Score (kontinuierlicher Rangwahrscheinlichkeitsscore) um 82% (48%) bei einer Vorlaufzeit von einem Tag.

Schließlich wurden regimebasierte Analysen durchgeführt, um zu beurteilen, ob Wetterregime die Niederschlagsmenge und die Fähigkeit der Vorhersage im Vergleich zur Referenzprognose EPC beeinflussen. Im Einzelnen wurden die Madden-Julian Oszillation (MJO), Kelvin-Welle, und der Indischer Ozeans Dipol (IOD) analysiert. Die Ergebnisse zeigen, dass MJO und Kelvin-Wellen die Niederschläge um 5-7 mmd^{-1} verändern. Die veränderung ist positiv (negativ) während der feuchten (trockenen) Phasen. Die veränderung findet durch die MJO hauptsächlich über Uganda und der ostafrikanischen Küste statt. Und bei der Kelvin-Welle über dem Kongobecken. Die Qualität der Vorhersagen verbessert sich während der feuchten Phasen bei der Erkennung von Niederschlag und der Genauigkeit der Niederschlagsmengen. Die IOD beeinflusst die Niederschlagsmenge und die Genauigkeit der Vorhersagen während der positiven, feuchten Phase positiv, während sie in der trockenen Phase das Gegenteil bewirkt. Die Arbeiten in dieser Arbeit haben den Grad der Genauigkeit der ECMWF-Vorhersagen, den Nutzen der Nachbearprozeddierung von Vorhersagen und die potenziellen Quellen der Vorhersagbarkeit aufgezeigt und dienen als Grundlage für die steigerung der Vorhersagequalität für die Region.

Preface

The PhD candidate confirms that the research presented in this thesis contains significant scientific contributions by himself. This thesis reuses material from the following publications:

> Ageet, S., A. H. Fink, M. Maranan, J. E. Diem, J. Hartter, A. L. Ssali, and P. Ayabagabo, 2022: Validation of satellite rainfall estimates over Equatorial East Africa. *J. Hydrometeor.*, **23**, 129–151, 10.1175/JHM-D-21-0145.1
>
> Ageet, S., A. H. Fink, M. Maranan, and B. Schulz, 2023: Predictability of rainfall over Equatorial East Africa in the ECMWF ensemble hindcasts on short to medium-range time scales (under review). *Weather and Forecasting.*

The research leading to these results has been accomplished with the support of a DAAD PhD Fellowship. The research proposal of this project was written by Simon Ageet and Andreas H. Fink. The analyses and results interpretation in Ageet et al. (2022) and Ageet et al. (2023), were solely performed by the candidate, who also wrote the text with advice from Andreas H. Fink and Marlon Maranan and comments from all co-authors during the manuscript preparation.
The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the publications of others.

Contents

AUC Area Under the Curve

AUCS Area Under the Curve Score

BID Bias in Detection

BS Brier Score

BSS Brier Skill Score

CCD Cold Cloud Duration

CHC Climate Hazards Center

CHIRP Climate Hazards Infrared Precipitation

CHIRPS Climate Hazards Infrared Precipitation with Stations

CORP Consistency, Optimality, Reproducibility and pool-adjacent-violators

CRPS Continuous Rank Probability Score

CRPSS Continuous Rank Probability Skill Score

CMIP Coupled Model Intercomparison Project

CP Convection Permitting

DJF December-January-February

DRC Democratic Republic of Congo

DSC Discrimination

EARS East African Rift System

EIG Eastward Inertio-Gravity

ENSO El Niño-Southern Oscillation

EMOS Ensemble Model Output Statistics

EEA Equatorial East Africa

EW Equatorial Waves

ER Equatorial Rossby

ECMWF European Centre for Medium-range Weather Forecasting

EPC Extended Probabilistic Climatology

GPD Generalised Pareto Distribution

GLOBE Global Land One-Km Base Elevation project

GPCC Global Precipitation Climatology Centre

GPCP Global Precipitation Climatology Project

GPM Global Precipitation Measurement

GPROF Goddard Profiling

GTS Global Telecommunication System

GDP Gross Domestic Product

HSS Heidke Skill Score

IDR Isotonic Distributional Regression

IG Inertio-gravity

IMERG Integrated Multi-satellite Retrieval for GPM

IOD Indian Ocean Dipole

IPCC Intergovernmental Panel on Climate Change

ITCZ Inter-Tropical Convergence Zone

JJA June-July-August

JJAS June-July-August-September

KASS-D Karlsruhe African Surface Station-Database

MAM March-April-May

MAE Mean Absolute Error

ME Mean Error

MCB Miscalibration

MJO Madden-Julian Oscillation

MRG Mixed Rossby-Gravity

MS Multisatellite

MSWEP Multi-Source Weighted-Ensemble Precipitation

NCEP National Center for Environmental Prediction

NMHS National Meteorological and Hydrological Service

NOAA National Oceanic and Atmospheric Administration

NWP Numerical Weather Prediction

OCHA The Centre for Humanitarian Data

OLR Outgoing Longwave Radiation

PAV Pool-Adjacent-Violators

PB Percent Bias

PIT Probability Integral Transform

ROC Receiver Operating Characteristic

POD Probability of Detection

POFA Probability of False Alarm

PoP Probability of Precipitation

POT Peak-Over-Threshold

PMW Passive Microwave

QQ Quantile-Quantile

RGs Rain Gauges

RG Rain Gauge

RMM Real-time Multi-variate MJO

RMSE Root Mean Square Error

ROC Relative Operating Characteristic

S2S Sub-seasonal to Seasonal

SOFF Systematic Observation Financing Facility

SOI Southern Oscillation Index

SON September-October-November

SOND September-October-November-December

SRE Satellite Rainfall Estimate

SST Sea Surface Temperature

TIR Thermal Infrared

TMPA TRMM Multi-satellite Precipitation Analysis

TRMM Tropical Rainfall Measuring Mission

UBOS Uganda Bureau of Statistics

UKMO United Kingdom Met. Office

UNC Uncertainity

UNEP United Nations Environmental Program

WIGOS WMO Integrate Global Observing System

WMO World Meteorological Organisation

WMRG Westward mixed Rossby-gravity

1. Introduction

The countries in Equatorial East Africa (EEA; Fig. 1.1) are heavily reliant on rain-fed agriculture. A large portion of these countries' gross domestic product (GDP) comes from agriculture with Burundi (Democratic Republic of Congo, DRC) having the highest (lowest) percentage, 39.5% (19.7%) compared to a world average of 6.4% based on 2017 estimates (CIA World Factbook 2023). This makes the region one of the world's most vulnerable to weather and climate related extremes, mainly in the form of floods and droughts. The IPCC (2022) estimates the region's death rate due to these disasters is 15-fold more than that of the less vulnerable regions of the world, e.g., western Europe. Currently, many countries in EEA are experiencing the longest and most severe drought, leaving about 70 million people at risk of starvation and death (Toreti et al., 2022). On the other hand, the most flood-prone regions of EEA have seen an increase in death and displacement of people due to floods and landslides. The Centre for Humanitarian Data (OCHA), estimates that between December 2019 and January 2020 alone, about 3.4 million people were affected in the region (OCHA, 2020). The number of people affected will grow in the future given the rapidly increasing population of EEA. For example, Uganda's population alone is projected to almost double to about 70.5 million by 2040 relative to the 2020 estimate (UBOS 2020). One important mitigation against weather and climate related disasters is the provision of early warnings informed by accurate weather and climate forecasts. Unfortunately, the potential of these forecasts to save life and property has not yet been realized for EEA and Africa at large (Youds et al. 2021).

One of the main reasons for the unrealised potential is the lack of skill of weather and climate forecasts in the region. Vogel et al. (2018) demonstrated that, even after postprocessing, daily rainfall forecasts from the European Centre for Medium-range Weather Forecasts (ECMWF) and National Oceanic and Atmospheric Administratio (NOAA) over Sub-Saharan Africa hardly beat climatology as a reference forecast. Haiden et al. (2012) showed that a forecast with 1-day lead-time in the tropics (where most of Africa lies) is equivalent to a 6-day lead-time forecast in the extra-tropics. These deficiencies of the models replicating the amounts of rainfall in the tropics have been partly attributed to inadequate parameterization of convection (e.g., Marsham et al.,

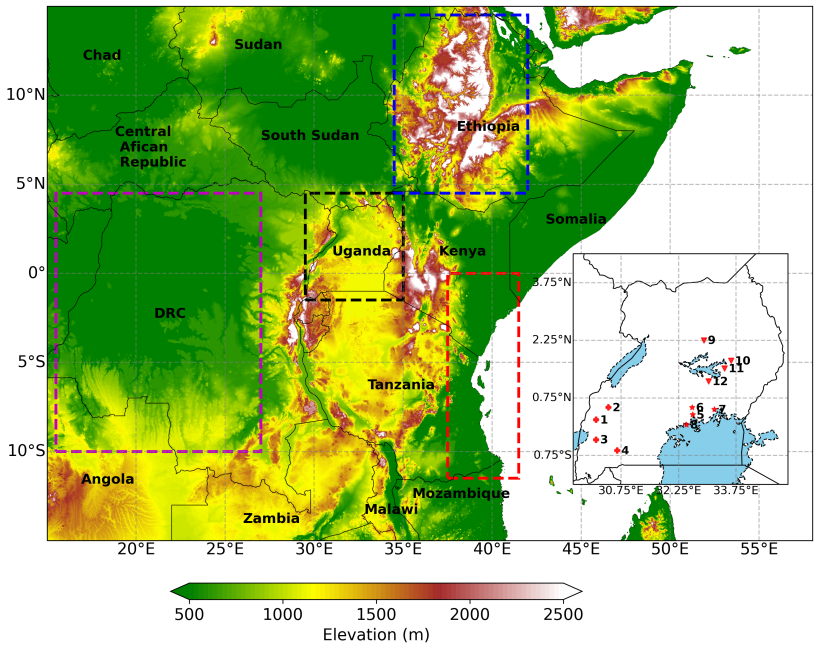


Figure 1.1: Study area with the four regions; Uganda (black), Congo Basin (magenta), Ethiopian Highlands (blue) and East African coast (red) demarcated. The shading show the highly varying elevation of the study area as given by the Global Land One-Km Base Elevation project (GLOBE). The map inset has the location of the 12 synoptic stations, grouped into four depending on the nearest 1.5° grid they fall in. These stations are representative of western- (1–4), central- (5–8) and eastern-Uganda (9–12)

2013; Birch et al., 2014; Vogel et al., 2018). Macleod et al. (2021) noted that using these poor forecast to make decision leads to wrong actions almost 50% in Kenya.

Ongoing research and increasing computational resources, has improved the representation of climate drivers, e.g., Madden-Julian Oscillation (MJO), El Niño-Southern Oscillation (ENSO), Indian Ocean Dipole (IOD), among others, in forecast models, leading to improved predictability in the region, especially at sub-seasonal time scales (de Andrade et al., 2021; Vitart et al., 2017; Li and Robertson, 2015). For EEA, studies have shown that most of these drivers, e.g., MJO (Pohl and Camberlin, 2006a; Specq and Batté, 2022) and IOD (Black et al., 2003; Wainwright et al., 2021), occur over EEA, suggesting the weather may be more predictable in the region, since these phenomenon are often well forecast in global models. However, these gains are not homogeneous over the globe for a number of reasons, e.g., difference in geography (e.g., the complex topography of EEA in Fig. 1.1), climate, and quality of initial conditions (or observations). Therefore, there is need to verify the skill of forecasts and calibrate for any bias and dispersion errors.

A number of studies have verified different (re)forecasts over EEA (e.g., de Andrade et al., 2021; Endris et al., 2021; Macleod et al., 2021; Stellingwerf et al., 2021). Reforecasts or

hindcasts are forecast with initialisation and prediction dates in the past but usually generated using the current operational forecast system. The verification studies generally suggest that the forecasts over EEA are skillful at lead times of weeks to months. Specifically, they found that; a) the reforecasts are more skillful over East Africa compared to those of other regions in Africa (de Andrade et al., 2021), b) ECMWF reforecasts generally performed best compared to reforecasts from other centres (de Andrade et al., 2021; Endris et al., 2021), and c) the performance of the reforecasts varies spatio-temporally. At daily time scales, Stellingwerf et al. (2021) also showed that the ECMWF model performed best.

Although there is promising skill in the forecasts over EEA, it was noted earlier that these forecasts have errors. For probabilistic forecasts, these are normally systematic errors in form of biases and dispersion errors. These errors are nowadays remedied by statistical postprocessing, which relies on past forecast-observation pairs to model the distribution of rainfall and hence, correct the numerical model outputs. This technique has been shown to work well, for example, Vogel et al. (2020) showed that, after postprocessing using Ensemble Model Output Statistics (EMOS; Gneiting et al., 2005), the forecast improved by an average of about 35%. Similar improvements were seen by Stellingwerf et al. (2021), who showed that bias correction using quantile-to-quantile mapping improved the skill of forecasts over Ethiopia.

The second cause of poor forecast skill in Africa is the low quality of observations. Žagar (2017) noted that the quality or low number of observations (initial conditions) going into the model determine its performance. However, in most African countries, data from RGs, often considered as reference data, are spatially and temporally sparse, temporally inconsistent, and in some cases of low-quality (e.g., Diem et al., 2014; Monsieurs et al., 2018). In fact, the station network coverage across Africa has been shrinking over the years (e.g., Asadullah et al., 2008; Dinku, 2019). Therefore, over Africa, hardly any RG data is assimilated into the forecast systems.
Satellite rainfall estimates (SREs) have filled the spatio-temporal data gaps. However, they are not without shortcomings. For example, IR-based techniques rely on a cloud top temperature-precipitation relationship, which is often unsuitable for non-convective precipitation (Kidd and Huffman, 2011) or non-precipitating cold cirrus shields (Young et al., 2014). On the other hand, microwave-based satellites have been found to struggle with retrieving rain from warm clouds (Dinku et al. 2010b; Monsieurs et al. 2018; Maranan et al. 2020) and to overestimate rainfall in places where convective rainfall dominates (Tian et al., 2009). Therefore, the SRE products used for validation of forecasts must be chosen carefully, and validated prior to their use.

From the above discussion, its clear that EEA needs accurate weather and climate information. One key pillar in the improvement of forecasts is validation. Besides assessing the skill of forecasts, validation also enables model development and discovery

of new sources of predictability. Most of the forecast validation studies done in this region either considered longer temporal aggregations, did not postprocess the forecasts (de Andrade et al., 2021; Endris et al., 2021), or used different post processing approaches Stellingwerf et al. (2021). Therefore, this work fills the existing gap by validating raw reforecast from ECMWF from short to medium-range time scales (i.e., 1–14 days ahead) for 24-hour, 48-hour and 120-hour accumulations. The raw reforecast will also be postprocessed and their skill evaluated. Additionally, the role of climate drivers in forecasting will be explored in a regime-dependent approach. Before carrying out the validation of the forecasts, identification of the right observation data sets is necessary. Therefore, the major aims of this dissertation are to:

1. Identify the most suitable satellite rainfall estimates for validating the forecasts over EEA.
2. Assess the skill of global ensemble forecast products over EEA.
3. Assess potential sources of rainfall predictability in EEA.

Chapter two provides a review of literature on the climate and weather of EEA, focusing on rainfall and its variability. Focus is put on the local and global climate drivers that bring about rainfall variability and the skill of global forecasts in this region. Chapter three expounds on the research goals, providing the research questions to achieve the goals. Next, results are presented starting with results of the validation of SREs in Chapter four. The best performing product at short time steps (i.e., daily to decadal accumulations) was used in the rest of the PhD project. In chapter five, results of the analysis of the skill of forecasts are presented, followed by regime-dependant skill analysis in Chapter six. Following the results chapters, the conclusions and outlook of this thesis are given in chapters seven and eight, respectively.

2. Background and literature review

In this chapter, the weather and climate of EEA and factors influencing it are discussed. The first section deals with the topographical features of the regions. The mean state and variability of the weather and climate resulting from the prevailing conditions of the regional and global climate drivers is presented in the second section. In the third section, the individual drivers are discussed, focusing on how they influence the weather and climate and their influence on the performance of global weather and climate models. Finally, in the fourth section, the historical and current status of forecasting in the region is explored. Additionally, factors affecting the quality and uptake of weather and climate forecasts are briefly introduced.

2.1 Topography of EEA

EEA has a complex topography (cf Fig.1.1); the region is dominated by the East African Rift System (EARS) of the oldest and most well-defined rift valleys in the world. The EARS starts from the Afar region in Ethiopia (Ethiopian rift) and splits into the western branch in the Lake Albert region (the Albertine rift) and the eastern branch which approximately bisects Kenya in a north-south line (Morley et al., 1999). The region has some of the highest landforms on the continent; the Ethiopian highlands with heights ranging between 1500 and 4550 m above sea level, the East African highlands where mountains like, Kenya (5199 m), Kilimanjaro (5895 m), Elgon (4321 m), Rwenzori (5109 m) and the Virunga ranges (3058–4507 m) are found. On the other hand, some regions, for example, the East African coast, Congo basin and South Sudan regions are low-lying, being as low as 0 m above sea level. The region is also home to the largest in-land water body in Africa, Lake Victoria, which spans three countries (i.e., Uganda, Kenya and Tanzania). Given its shear size of ~69,000 square kilometers (UNEP, 2006), it is the most significant lake both in terms of socio-economic and climatic significance. Other major lakes in the region include, Albert, Edward, Kyogo, Kivu, Mweru, Tanganyika, Malawi, Rukwa and Turkana. These lakes form what is referred to as the "African Great Lakes" or the "Rift valley lakes".

The influence of topography on rainfall, and hence climate, of a region is well known. In East Africa, Oettli and Camberlin (2005) showed that slope, exposure and elevation of

the topography explained the variation of rainfall in the region. Based on the RegCM4 climate model, Ogwang et al. (2014) showed that a 25% decrease in elevation resulted in a 19% reduction in October-November-December rainfall in East Africa. Lake Victoria influences the local climate of the surrounding regions via its dominant land-lake breezes. Given its direct influence on weather and climate, it can be expected that the skill of forecasts is affected by the topography of a region. For example, over Lake Victoria, Woodhams et al. (2018) showed that better representation of land-lake breeze dynamics in the convection-permitting (CP) Met Office operational model resulted in a 20% increase in hit rates for forecast of extreme precipitation relative to the global Met Office model with the usual parameterized convection. Similar results were shown by Cafaro et al. (2021). This shows that models that can effectively represent the localised convection over the lake show superior skill. Models also seem to better represent topographically-triggered convection over EEA. Moreover, over the East African highlands, convection is often linked to large-scale triggers, such as the MJO (Pohl and Camberlin, 2006a) which have higher predictability (e.g., de Andrade et al., 2021; Vitart et al., 2017). On the other-hand, over flat low-lying areas and where convective rainfall dominates, the rainfall is more stochastic (Satheesh et al., 2023), and therefore, models fail to correctly represent this convection.

2.2 Climatology of EEA

2.2.1 Mean state

The climate of EEA varies significantly, both in space and time (e.g., Peel et al., 2007; Nicholson, 2017; Seregina et al., 2018; Beck et al., 2018; Dunning et al., 2018). The Koppen-Geiger climate classification from Peel et al. (2007) shows EEA has at least five different climates, including tropical Savannah, arid desert and arid steppe being the dominant types (Fig. 2.1). According to Nicholson (2017), these different climates are as a result of; a) local topographical feature (e.g., Turkana channel, East African and Ethiopian highlands, Lake Victoria and Red Sea trench, b) regional circulations features (e.g., tropical easterly jet, low level westerlies, localised convergence, and Turkana jet), c) remote forcing (e.g., Walker circulation, IOD, ENSO, MJO and global SSTs and d) coastal influences (e.g., frictional uplift, Somali jet, coastal SSTs and sea breeze). Some of these factors will be explained in the coming sections. EEA is heavily impacted by severe drought, owing to the arid and semi-arid climate. For example, the region has recently been plagued by the longest and most severe drought in recent times, stretching for five rainy season between 2020 and 2022 (CHC, 2023). The largely dry climate of eastern African is an anomaly for a region close to the equator (Griffiths and Landsberg, 1972). The subsiding arm of the Walker circulation, one of the major circulation features in the region may partly explain this dryness (see Walker circulation explanation in section 2.3.2). Nicholson (1996) suggests that thermally stable monsoon flows and stable

air aloft are possible reasons. On average, most areas receive between 800–1200 mm of rainfall annually (Nicholson, 2017).

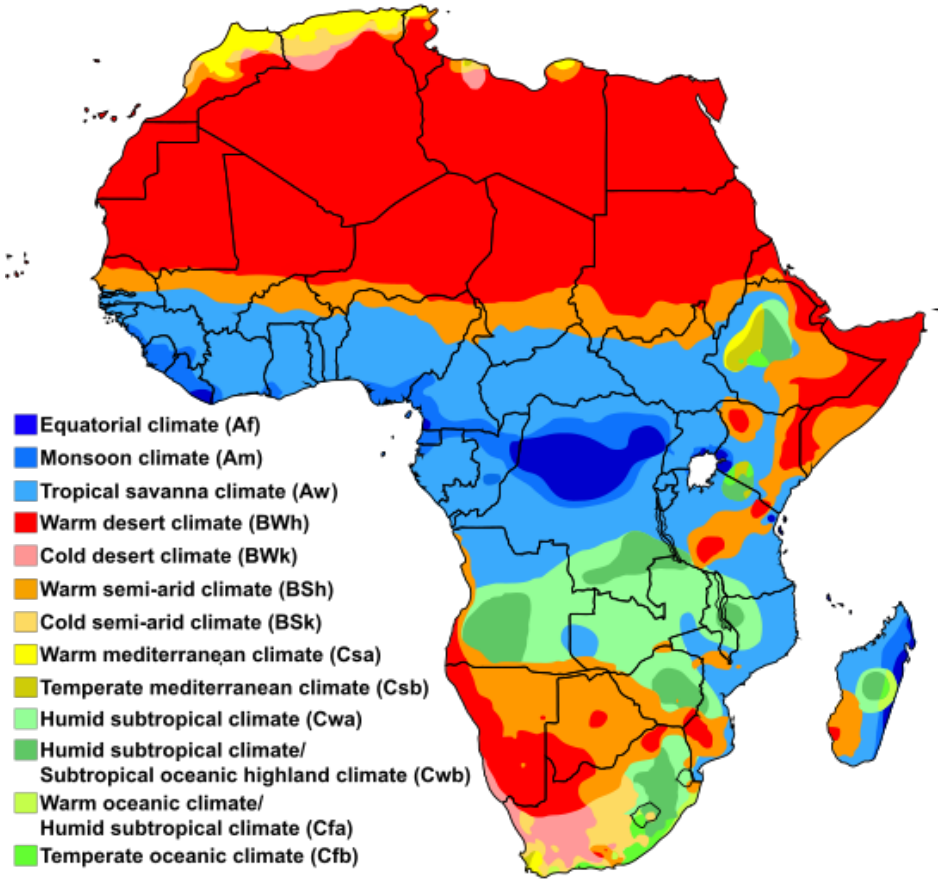


Figure 2.1: The climate over Africa according to the Koppen-Geiger classification. Taken from Peel et al. (2007).

2.2.2 Seasonal rainfall cycle over EEA

Rainfall in EEA is seasonal with the number of rainy seasons varying from one to three depending on location. Larisa et al. (2021) delineated the regions receiving unimodal, bimodal, and trimodal rainfall regimes (Fig. 2.2). The unimodal regime dominates the northern part of region in the boreal summer from June to September(JJAS) and the southern part during December to February (DJF) season. Although some regions have a trimodal rainfall cycle, the third rainy season is mostly unstable. The occurrence of these seasonal cycles is strongly influenced by the migration of the Inter-Tropical Convergence Zone (ITCZ) through the year (e.g., Dunning et al., 2016; Nicholson, 018a; Seregina et al., 2018). Because it is an area where the trade winds converge, the ITCZ is generally an area with enhanced convective activity. In the boreal summer, when it is located in the northern hemisphere, the rainfall peak is in the northern part of EEA, hence the unimodal rainfall peaking at this time of the year. The reverse is true during the boreal winter, when the mean location of the ITCZ is in the southern hemisphere,

single rainfall peak in the southern parts of EEA. Because the ITCZ crosses some locations twice, during its northward and southward legs, these locations receive two rainfall peaks, hence the bimodal rainfall cycles. Because of the gradual migration of the sun, the length, start and end of seasons is different across EEA.

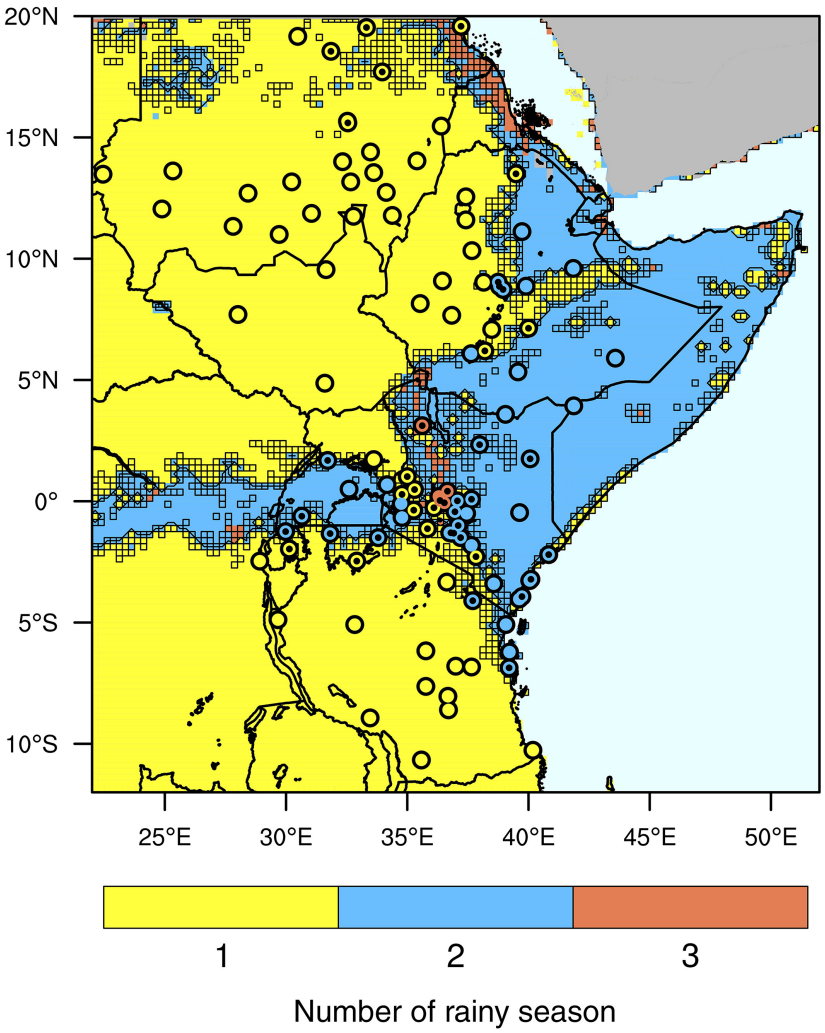


Figure 2.2: Number of rainfall seasons across EEA; yellow for unimodal, blue for bimodal and orange for trimodal rainfall seasons based on CHIRPS and rain gauge data (circles). Grids(gauges) with unstable number of seasons are hatched(black-dotted). Taken from Seregina et al. (2018), Figure 6, © 2018 John Wiley & Sons, Ltd.

Fig. 2.3 shows the mean daily rainfall in each month over EEA with a gradual shift in the location of maximum convection, following the ITCZ movement. Starting in January, the rainfall maximum moves north-wards until July/August, when it is at the northernmost location, e.g., over the Ethiopian highlands, South Sudan and Sudan regions. The convective regions start to reverse direction in September, moving southwards. One of the major circulations associated with this annual cycle is the trade winds, which occur as southwest (SW) monsoons in the boreal summer and northeast (NE) monsoons during the boreal winter. These winds are shown by the 850 hPa vectors from ERA5 in Fig.

2.3. During the summer, the sun is mainly over the northern hemisphere which creates a low pressure region. Onshore southeasterly winds therefore move equator-wards before bending to become southwesterlies after crossing the equator. This flow pattern is associated with the Somali low level jet (Findlater, 1969), which is located between 1200–2400 m above sea level with a maximum speed of 25–50 ms^{-1}. The jet starts from the southern hemisphere as southeasterly flow and becomes southwesterly upon crossing the equator. The Ethiopian highlands play a big role in confining the jet along the Indian Ocean coast as it moves towards India where most of the moisture is dumped, causing the intense "monsoon rains" over the Indian peninsular. In September, the winds start to reverse in direction becoming northeasterly and then northwesterly after crossing the equator.

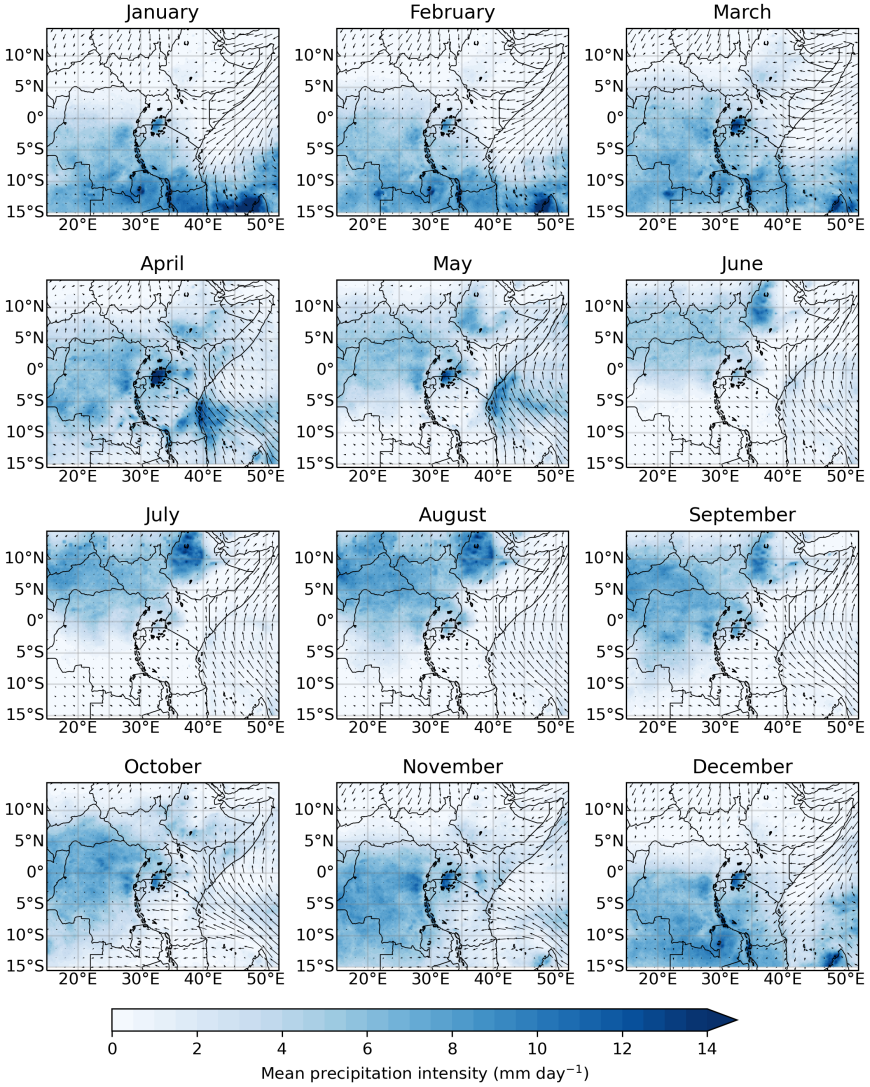


Figure 2.3: Mean daily precipitation in eaach month of the year over EEA based on IMERG-F precipitation for the period 2001–2018. The 10 m wind (vectors) based on ERA5 Hersbach et al. (2020) for the period 1979–2020.

2.3 Modes of Rainfall variability over EEA

Rainfall over EEA is highly variable. The variation happens within a day (diurnal variation), within a season (intraseasonal variation), among seasons (seasonal variation) and between years (interannual variation). As explained earlier, the number of rainy seasons varies between one and three depending on the location. Additionally, the amount of rainfall, onset, cessation and length of these seasons are spatio-temporally variable (e.g., Nicholson, 2017; Seregina et al., 2018; Dunning et al., 2018). Within specific locations and seasons, the rainfall received in different years substantially varies. For example, Palmer et al. (2023) estimates that rainfall on average varies by about 30% (60%) during the long (short) rain season in Eastern Africa Sahelian region. They also noted that a maximum deviation of 355% from the long-term mean was reached in 1997. These results are in line with those of Ogallo (1989) and Nicholson (1996) who found that the short rain season has a greater variability compared to the long rain season. Even the diurnal cycle of daily rainfall is different across the region. Camberlin et al. (2018) showed that the diurnal cycle of rainfall is quite stable over EEA, with morning and afternoon maxima along the East African Coast (EAC) and over land, respectively. The quite stable states have been shown to be modulated on intraseasonal and interannual time scales by MJO and IOD (Camberlin et al., 019b). The physical processes that drive the rainfall variability, commonly termed 'climate drivers' are explained in the following subsections.

2.3.1 Madden-Julian Oscillation

The largest mode of intraseasonal rainfall variation is the Madden-Julian Oscillation (MJO; Madden and Julian, 1971, 1972; Zhang, 2005). This planetary-scale disturbance which propagates eastward at ~5 ms^{-1} is associated with convection and traverses the tropics in 30–90 days (Zhang, 2005). The phenomenon was first discovered by Madden and Julian (1971) who noticed opposite oscillations in zonal wind anomalies in the lower (850hPa) and upper (150hPa) atmosphere. These oscillations were later shown to be associated with enhanced convection where the zonal winds at low levels converged and dry conditions where they were divergent. Because the MJO is active in different locations at different times and brings either enhanced or suppressed convection, an index was proposed to differentiate the location and associated effect of the MJO. The Real-time Multi-variate MJO (RMM; Wheeler and Hendon, 2004) index, classifies the MJO into eight phases depending on the location of enhanced convective activity (Fig. 2.4). The convection is enhanced over the western hemisphere and Africa in phases 1 and 8, Indian Ocean in phases 2 and 3, Maritime continent in phases 4 and 5, and the western Pacific in phases 6 and 7, all the time propagating eastwards.

Several studies have shown that the MJO substantially modulates rainfall over East Africa (e.g., Pohl and Camberlin, 2006a,b; Berhane and Zaitchik, 2014; Camberlin et al.,

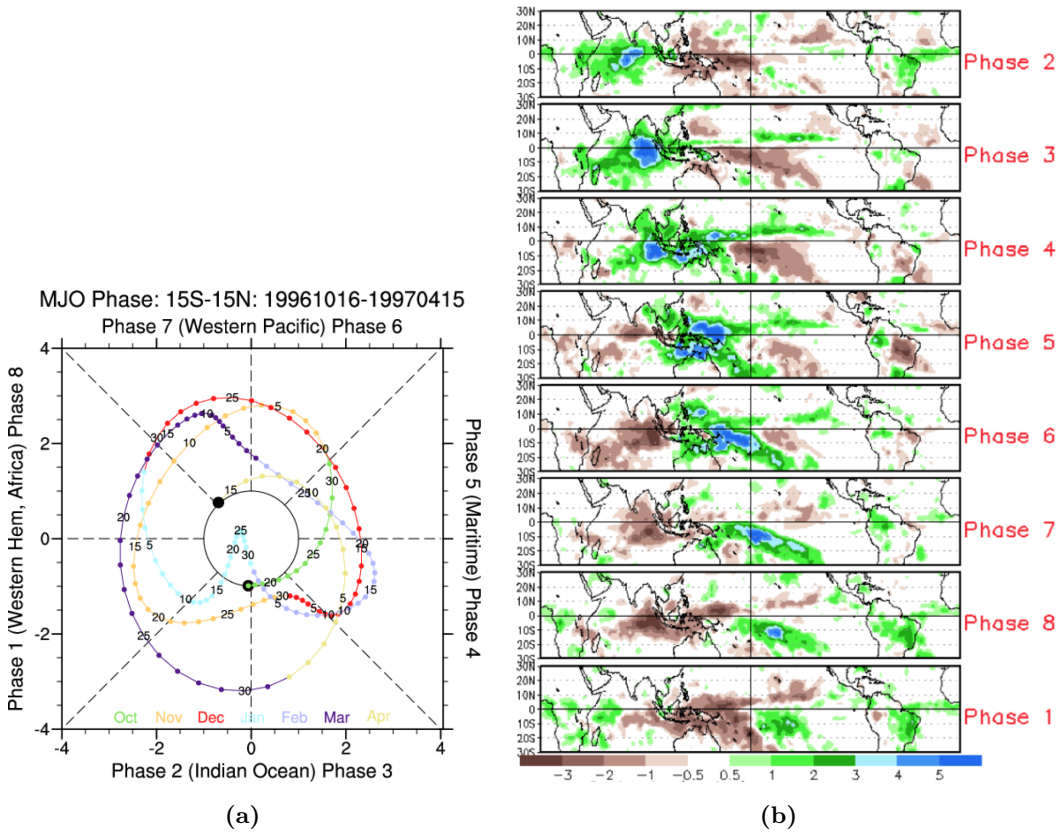


Figure 2.4: (a) The RMM phase diagram showing the state of the MJO between Nov 2022 and February 2023. Source: Climate Data Guide; D. Shea. (b) Rainfall anomaly during MJO events between 1979-2012 for November-March for each of the eight phases. The green (brown) shading denotes above-average (below-average) rainfall. The eastward propagation of the MJO can be infered from the eastward shift in the blue shading. The phases are also labeled logically following the eastward propagation of the rainfall modulation. Source: NOAA

2018; Specq and Batté, 2022; Vashisht and Zaitchik, 2022) and is the leading driver of rainfall variability on a sub-seasonal scale (e.g., Specq and Batté, 2022; Palmer et al., 2023). Berhane and Zaitchik (2014) and Vashisht and Zaitchik (2022) showed that the most affected months are March and May in the long wet season and November–December in the short wet season, while in April, the MJO is obscured. The MJO phases associated with rainfall in East Africa, especially over the highland regions, are 2–4 (Pohl and Camberlin, 2006a; Berhane and Zaitchik, 2014). These phases are actually coupled with with heavy convection over the Indian Ocean (cf. Fig. 2.4). The rainfall over East Africa in these phases is a result of the large scale convection envelopes over the Indian Ocean resulting to westerly moisture advection. Pohl and Camberlin (2006a) adds that phases 6–8, which lead to suppressed convection over the African/Indian Ocean regions, lead to enhanced convection in the low-lying eastern and coastal regions of EEA, owing to moisture advection from the Indian Ocean. Active MJO amplitudes bring about either, wetter or drier than normal spells that can last up to three weeks

(Vellinga and Milton, 2018). The strong wet spells are associated with early seasonal rainfall onsets and extreme precipitation accumulations and this may account for up to 44% of seasonal rainfall variation (Pohl and Camberlin, 2006b). Overall, during a season, Vellinga and Milton (2018) note that the net effect of strong MJO activity is increased rainfall owing to the fact that the relationship between MJO amplitude and rainfall amounts is non-linear, as rain amounts are capped at zero.

The MJO has been linked to increased predictability of weather, leading to better quality forecasts, especially at the subseasonal scale (e.g., Vitart et al., 2017; de Andrade et al., 2021). de Andrade et al. (2021) showed that, for EEA, the MJO has the biggest influence on the skill of the ECMWF reforecasts. They showed that the skill in weeks 3–4 improved most when the MJO-related rainfall variability was added in comparison to other climate drivers. Additionally, Specq and Batté (2022) showed that models do convert MJO information into better forecasts by increasing the hit rate of events over Uganda. However, they also noted an increase in false alarm rate and misplaced positions of the MJO signal. Given that EEA is one of the regions where MJO signals are strongest, and that models have become better in predicting MJO signals (Vitart et al., 2017), it is expected that the skill of forecasts is enhanced in this region compared to other regions of Africa.

2.3.2 El Niño-Southern Oscillation and Indian Ocean Dipole

El Niño-Southern Oscillation (ENSO) is a periodic (3–7 years) mode of climate variability. It is composed of two parts; a) fluctuation of sea surface temperatures (SSTs) of the Pacific Ocean (El Niño), and b) oscillation of pressure gradient between eastern (Tahiti) and western (Darwin) Pacific Ocean (Southern Oscillation). These two components of ENSO were discovered separately. Walker (1925) noticed that the pressure gradient between the eastern and western Pacific fluctuated back and forth on a large scale. This fluctuation was later defined using an index, the Southern Oscillation Index (SOI; Chen, 1982), derived by standardizing the pressure difference between Darwin and Tahiti. The negative (positive) phase of the SOI happens when the pressure at Darwin (Tahiti) is above (below) normal. During a prolonged negative (positive) phase of SOI, the water over the eastern pacific is abnormally warm (cold), what is now known as El Niño (La Niña). The abnormal warming of water was first discovered by fishermen off the coast of Peru who noticed that, every few years, the waters would warm up abnormally leading to reduced fish catch. They termed this El Niño, mean boy child as the phenomenon usually occurred around Christmas. They also noticed that in some years, the water was abnormally cold, which they termed La Niña (which translates to girl child). It was later discovered by Bjerknes (1966) that Oceanic and Atmospheric parts of ENSO were coupled. Therefore, ENSO is a coupled Ocean-Atmosphere phenomenon and the greatest mode of interannual climate variation.

ENSO can be in either the El Niño, La Niña or neutral phase (Fig. 2.5) with different implications on the weather in different locations. During normal conditions, SSTs in the tropical east Pacific are cooler than those in the tropical western Pacific. The resulting temperature gradient in the Ocean basin means cool air from the east blows toward the warmer west Pacific. The air picks up moisture and leads to convection over the Maritime Continent. This circulation is known as the Walker circulation, a term first introduced by Bjerknes in 1960 in honour of Sir Gilbert Walker (Bjerknes, 1969). As shown in the top panel of Fig. 2.5, this zonal circulation over the equatorial pacific is characterised by surface flow from the western Pacific to the warm waters of the central Pacific leading to the rising motion. At the upper level a returning, divergent flow completes the circulation. Over EEA, the walker circulation is convergent (divergent) at the upper (lower) level. The semi-aridity of most parts of EEA has been attributed to mainly divergent flow at the surface (e.g., Nicholson, 2017) and the relatively stable atmosphere (Yang et al., 2015). The importance of the Walker circulation is amplified when an El Niño-Southern Oscillation (ENSO) occurs as explained in the next paragraphs.

Because of the easterly flow over the tropical Pacific during normal phase of ENSO, warm water is pushed to the west and upwelling of cold water from deep in the ocean off the coast of Peru. This leads to further cooling of the water over the east and central Pacific. This feedback effect, known as the Bjerknes feedback, leads to the cool phase of ENSO, the La Niña (the bottom panel of Fig. 2.5). In this phase, the rising motion over the Maritime Continent is strengthened, while the opposite occurs over EEA. Strengthening of the subsiding arm of the Walker circulation over East Africa has been linked to drier conditions (Williams and Funk, 2011). Because of the pile up of water in the west Pacific, the sea level is higher here than in the eastern Pacific. This leads to weakening of the easterlies due to a higher pressure in the west. Consequently, the warmer water from the west flows down-slope towards the eastern Pacific, increasing the SSTs here. Warmer SSTs lead to rising motion which further lower the pressure, hence, more westerly flow and warming of the SSTs. Just as in the La Niña phase, the Bjerknes feedback keep the SSTs increasing. This results in anomalous warming of the eastern Pacific, leading to the warm phase of ENSO, El Niño (middle panel of Fig. 2.5). During this phase, the Walker circulation is shifted with reversed regions of subsidence and ascent. Over EEA, there is a raising arm of the circulation, leading to enhanced rainfall during this phase.

A related but different event, the Indian Ocean Diplole (IOD) occurs over the Indian Ocean. The IOD refers to the periodic oscillation of the SST anomaly between two areas (or poles, hence, dipole), the western (50°E–70°E and 10°S–10°N) and eastern (90°E–110°E and 10°S–10°N) Indian Ocean (Palmer et al., 2023). This interannual mode of variability was termed Indian Ocean dipole or Zonal mode (IOZM; Saji et al., 1999; Webster et al., 1999). The positive and negative phases of the IOD occur when a

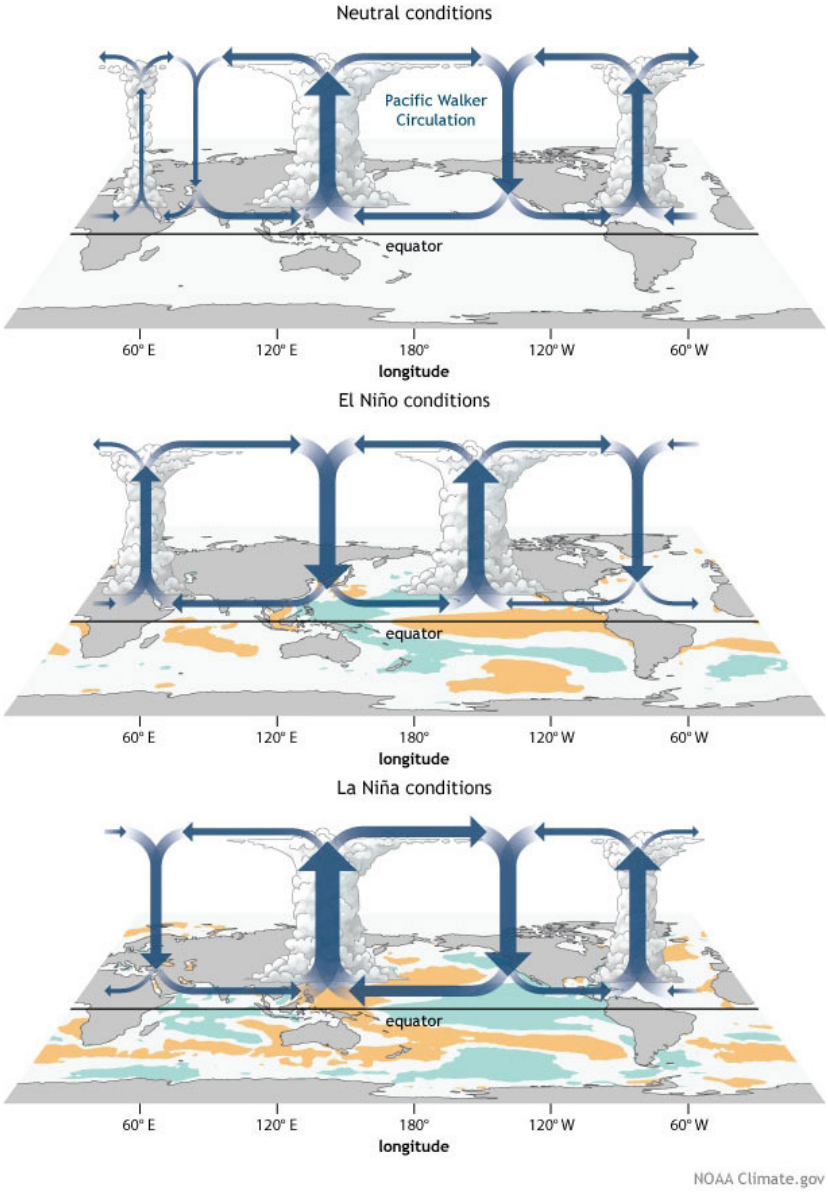
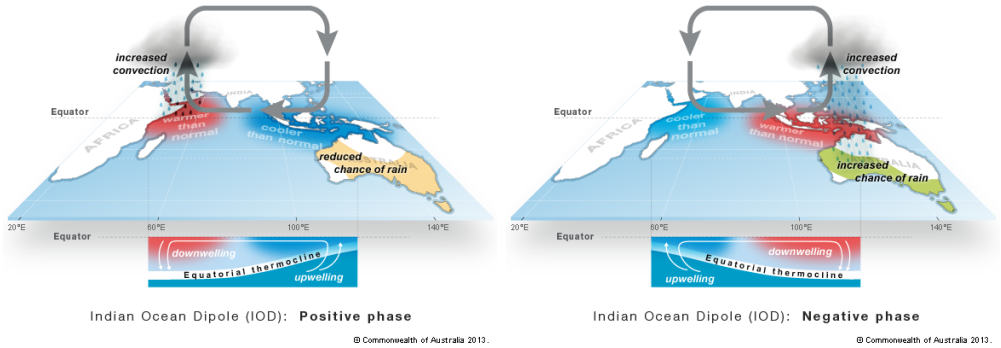


Figure 2.5: Mean SSTs (shading) and Walker circulation (arrows) during the Neutral (top), El Niño (middle) and La Niña phases of ENSO. Source: NOAA

temperature anomaly of 0.4°C above and below zero, respectively, is sustained for at least three months. Like ENSO, the IOD is a coupled atmosphere-Ocean phenomenon that happens over the Indian Ocean. The positive (negative) phase (Fig. 2.6) is characterized by warmer (cooler) SSTs in the western (eastern) Indian ocean, anomalous easterly (westerly) winds at the surface, leading to more (less) than normal rainfall over East Africa (western Indian Ocean). As stated in Palmer et al. (2023), 1997, 2006, 2012, 2015 and 2019 (1996, 1998, 2010 and 2016) were positive (negative) IOD years. The effect of the IOD rains is well documented. Hirons and Turner (2018) notes that the positive phase of IOD leads to enhanced rainfall in the short rain season due to anomalous low-level easterly wind flow across the equatorial Indian Ocean. The rainfall

accumulation can be up to 2-3 times the normal and lead to floods like seen in 2019, when one of the strongest IOD events on record occurred (Wainwright et al., 2021). On the other hand, the negative IOD phase substantially reduces the amount of rainfall during the short wet season (e.g., Black et al., 2003).



(a) **(b)**

Figure 2.6: Schematic of the IOD; a) positive phase and b) negative phase and their impact on the Walker circulation. The shading is the SST anomalies (red and dark-blue indicating warmer and cooler than normal) while the arrows indicate the state of the Walker circulation. The Walker circulation is completely reversed in the opposite phases Source: Australian Bureau of Meteorology; http://www.bom.gov.au/climate/iod/.

ENSO and IOD phenomena have similar dynamics, occur at the same time of year and their phases are the same when both occur together in a given year. Both modify the Walker circulation depending on the phase (cf. Figs. 2.5 and 2.6). Indeed, some people have questioned if ENSO and IOD really are different phenomenon (e.g., Marchant et al., 2007; Meyers et al., 2007). Whilst an El Niño can occur concurrently with a positive IOD event, as was the case in 1982 and 1997 leading to more intense anomalies and effects such as flooding (Webster et al., 1999; Wainwright et al., 2021), the latter also happens in neutral ENSO years, for example the very extreme IOD event of 1961 and 2019 (Saji et al., 1999; Wainwright et al., 2021). Indeed, Yamagata et al. (2004), suggested that only one in three IOD events are related with ENSO. Hence, IOD and ENSO should not be viewed as one and the same. The two are however interlinked, with IOD strongly modulating the impact of ENSO on rainfall (MacLeod et al., 2021). The combined occurrence and modulation leads to more extreme rainfall over EEA (Yamagata et al., 2004; Bahaga et al., 2014), as was the case in 1997 when more than double the normal climatological rainfall accumulation was seen in the short wet season (Wainwright et al., 2021). On the contrary, in 2015, when a strong El Niño occurred but with a weak IOD event, the rainfall anomaly was only about 50% above the climatology (Indeje et al., 2000). Ogallo et al. (1988) also showed that almost 50% of seasonal rainfall variability is due to ENSO. Overall, over EEA, ENSO and IOD are the leading drivers of interannual rainfall variability. It has been shown that the influence of ENSO influences both the short and long rains (Ogallo et al., 1988; Indeje et al., 2000; MacLeod et al., 2021),

while the effect of IOD is felt only in the short rain season as it start in September lasting up to January of the next year (Black et al., 2003; Shaaban and Roundy, 2017). During the short rain season, if an El Niño and/or positive IOD event occur, the Indian Ocean Walker cell is weakened (e.g., Nicholson, 2015; Jiang et al., 2021) enabling the anomalous warming of SSTs over the western Indian ocean with a corresponding cooling of SSTs over Southeast Asia. This leads to anomalous rising (sinking) of air over the East Africa/western Indian Ocean (eastern Indian Ocean) resulting to enhanced (reduced) rainfall over East Africa (Southeast Asia) (Palmer et al., 2023). Because the El Niño may be followed by a La Niña event, ENSO mostly negatively impacts the long rains. According to Funk et al. (2018), El Niño may be followed by warm western Pacific Ocean which strengthen the Indian Walker cell, hence, suppressing rain over East Africa.

Because of their slow evolution, ENSO and IOD are usually quite well forecast by most forecasting centres. Both numerical weather prediction and statistical forecasting techniques have been applied to forecasts of these phenomenon (e.g., Wu et al., 2009; Barnston et al., 2012; Lima et al., 2009; Wang et al., 2020; Zhao et al., 2020). The fact that most models incorporate atmosphere-ocean interaction improves the models performance (e.g., Wu et al., 2009; Barnston et al., 2012). Recently, machine learning approaches have been explored to improve ENSO forecasts (e.g., Haoyu et al., 2023; Ham et al., 2019; Liu et al., 2021).

2.3.3 Congo Air Mass

One of the major sources of moisture in Equatorial Africa is the Congo forest, the others being the Indian Ocean and Lake Victoria (Finney et al., 2019). A westerly flow from the Congo basin towards East Africa, referred to as the Congo air mass therefore brings moisture to East Africa. The occurrence of westerlies is linked to the MJO, specifically phases 2–4 (e.g., Pohl and Camberlin, 2006a; Hogan et al., 2014) suggesting that the rains may be due to the MJO. However, as explained in section 2.3.1, phases 2–4 of the MJO are associated with convection over Indian Ocean rather than over Africa, suggesting an indirect link (Finney et al., 2019). This link can be explained by the Matsuno-Gill response which suggests existence of anomalous westerlies to the west of the MJO core (Matsuno (1966); Gill (1980).The existence of the westerly flow and its rainfall modulation, especially over the East African highland has been suggested in past studies (e.g., Pohl and Camberlin, 2006a; Berhane and Zaitchik, 2014; Finney et al., 2019). Theses studies add that the coastal sides which lie on the leeward side are drier. Finney et al. (2019) attributed the intense rainfall of 2018 to the Congo airmass and suggest that up to 100% more rainfall can be recorded. However, they also noted that the number of days with absolute westerly flow is small, reaching a maximum of 11% of the days in July with two minima in April and November (see their Figure 3(c)). The short rain seasons has less westerly days compared to the long wet seasons. Generally,

on these days, the moisture flux and precipitation are enhanced over most parts of EEA (Fig. 2.7).

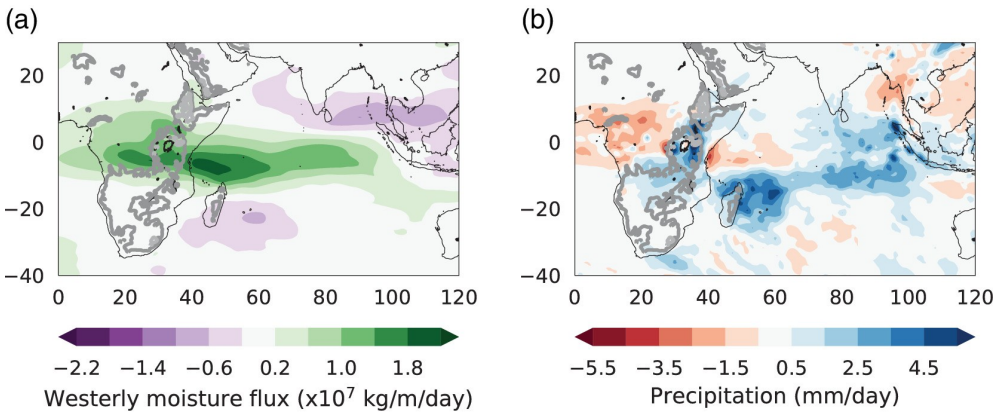


Figure 2.7: Mean difference of atmospheric state on days with westerly moisture flux compared to days with an easterly flux. ERA-Interim variables considered are: (a) column-integrated westerly moisture flux and (b) precipitation. Taken from Finney et al. (2019), Figure 7, © 2019 John Wiley & Sons, Ltd.

Because the westerlies explained above lead to variability of rainfall in EEA on subseasonal time scale, they influence predictability of the weather. Finney et al. (2019) and Walker et al. (2019) hypothesize that, since the short rain season has fewer westerly days, it has a better predictability on seasonal time scale compared to the long season with more westerly days. Finney et al. (2019) suggests that this is because the westerlies are not a well-known weather driver as they occur on short time scales and over synoptic scales. Although this explanation is contrary to the assertion that MJO has improved the predictability in this region (e.g., Vitart et al., 2017; Vellinga and Milton, 2018; de Andrade et al., 2021; Specq and Batté, 2022), the MJO forecasts are often not accurate, putting its locations in wrong places and with biased intensities (e.g., Specq and Batté, 2022).

2.3.4 Equatorial waves

The presence of propagating disturbances trapped to the equator, known as Equatorial waves (EWs), were long suspected before Matsuno's confirmation in 1966 (Ichiye, 1959; Stern, 1963; Matsuno, 1966). These disturbances, which are synoptic-to-planetary-scale (Knippertz et al., 2022), are often coupled with deep convection (Takayabu and Nitta, 1993; Takayabu, 1994; Wheeler and Kiladis, 1999) and therefore influence rainfall variability over the tropics. While these waves are briefly introduced in the next paragraph, a detailed description and derivation of EWs can be found in (e.g., Matsuno, 1966; Wheeler and Kiladis, 1999; Kiladis et al., 2009). In summary, the different waves are based on the solutions to the shallow water equations solutions solved in Matsuno (1966). With the assumptions of no background flow and that the Coriolis parameter

f varies linearly with latitude y i.e., the β-plane approximation ($f = \beta y$), Matsuno showed that the solutions for zonally propagating waves satisfy the relationship

$$K\frac{\sqrt{gh_e}}{\beta}(\frac{\omega^2}{gh_e} - k^2 - \frac{k}{\omega}\beta) = 2n+1, n = 0, 1, 2, ..., \tag{2.1}$$

where; β and k are the frequency and wavenumber of the wave, respectively, g is gravitational acceleration, h_e an equivalent depth and n is the number of meridional modes (Kiladis 2009). The waves that satisfy the relationship are:

1. Equatorial Rossby (ER) waves with the dispersion relation

$$\omega_{ER} \approx \frac{-\beta k}{k^2 + (2n+1)\beta/\sqrt{gh_e}}. \tag{2.2}$$

   This is a westward moving wave acting on time scales of 30–100 days and is obtained by setting $n = 1$ in Eq. 2.1. ER waves modulate precipitation, with the shorter modes impacting the S2S (10–30 days) time scale Schlüter (019a). Although the impact of ER on convection variability is less than that due to the MJO and Kelvin wave, at off-equatorial locations like the Philippines and Northern Australia its impact is highest (Wheeler and Kiladis, 1999).

2. Mixed Rossby-Gravity (MRG) waves which have the dispersion relation

$$\omega_{MRG} = k\sqrt{gh_e}\frac{1}{2} \pm \frac{1}{2}(1 + \frac{4\beta}{k^2\sqrt{gh_e}})^{1/2}, \tag{2.3}$$

   obtained when $n = 0$ in Eq. (2.1). They were first detected as low frequency waves oscillations propagating westward in the lower stratosphere (Yanai and Maruyama, 1966).

3. Inertio-gravity (IG) waves which are eastward MRG waves, sometimes referred to as eastward inertio-gravity (EIG) waves with the dispersion relation

$$\omega_{IG} \approx \pm[(2n+1)\beta\sqrt{gh_e} + k^2 gh_e]^{1/2}. \tag{2.4}$$

4. Kelvin waves with the dispersion relation

$$\omega_K elvin = \sqrt{gh_e k} \tag{2.5}$$

   obtained when $n = -1$ in Eq. (2.1) and are eastward propagating. This wave is almost equal in importance as the MJO in terms of causing convective variability. The wave is faster, occurs more globally in longitude and its more confined and centred on the equator (Wheeler and Kiladis, 1999).

As summarised in Knippertz et al. (2022), there are several methods of identifying EWs in observations. They can be categorized into two major groups, namely; a) broad filter windows that consider space only, such as, the 3D spatial projection using Hough functions, 2D spatial projection using parabolic functions and 2D spatial projection

using time-extended empirical orthogonal functions, and b) narrow filter windows which consider time and space, like the frequency-wavenumber filtering using parabolic functions, frequency-wavenumber filtering unsing Fast-Fourier transform and frequency-wavenumber filtering using wavelets. For a detailed discussion and comparison of these methods, the reader is referred to Knippertz et al. (2022). In this thesis, only the Fast-Fourier approach following Wheeler and Kiladis (1999) is introduced and applied in chapter 6. In this approach, 2D horizontal fields of de-seasonalized anomalies, for example precipitation, outgoing long wave radiation (OLR) or some other suitable weather variable is time-space filtered. The individual waves introduced above are identified based on set $k - \omega$ power spectra thresholds. Schlüter (019a) summarised the characteristic of the waves as shown in the Table 2.1. Following the thresholds above,

Table 2.1: Characteristics of the different EWs introduced above. The period T, the planetary zonal wave number k, and the equivalent depth h_e can be used to filter the waves. Adapted from Schlüter (019a) (Table 4.1). Used with permission

Acronym	Wave	Direction	T (days)	k	h (m)	Source
MJO	Madden-Julian Oscillation	Eastward	30 – 96	0 – 9	-	Roundy and Frank (2004)
ER	Equatorial Rossby wave	Westward	9 – 72	1 – 10	1 – 90	Kiladis et al. (2009)
MRG	Mixed Rossby-gravity wave	Westward	3–8	1 – 10	8 – 90	Wheeler and Kiladis (1999)
Kelvin	Kelvin wave	Eastward	2.5 – 20	1 – 14	8 – 90	Wheeler and Kiladis (1999)
EIG	Eastward inertio-gravity wave	Eastward	1–5	0 – 14	12 – 50	Yasunaga and Mapes (2012)

the popular dispersion relationship in the zonal $k - \omega$ domain for the waves can be plotted as shown in Figure 2.8 taken from Wheeler and Kiladis (1999).

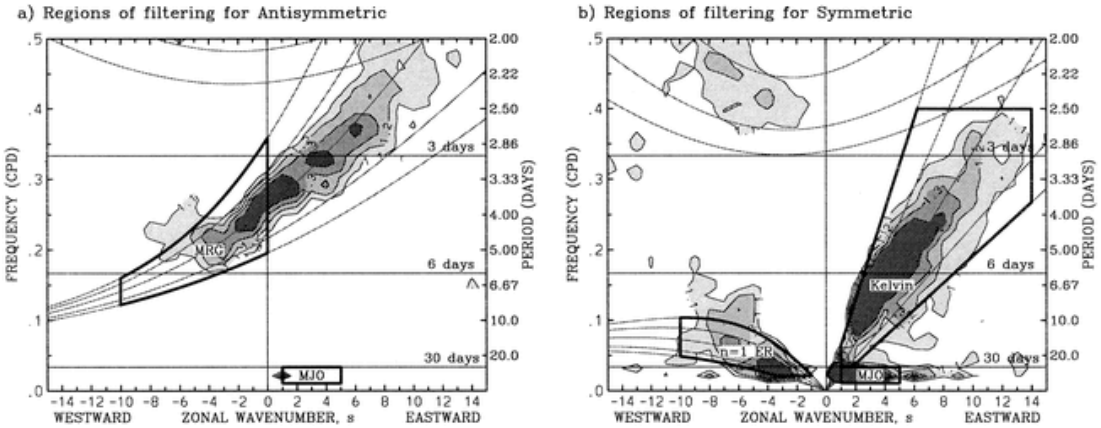


Figure 2.8: a) Antisymmetric and b) symmetric power spectrum of OLR taken between 15°N and 15°S. Thick black boxes are the regions of the wavenumber-frequency domain used for filtering of the OLR dataset to retrieve the longitude-time information of the convectively coupled tropical waves. The thin lines are the various equatorial wave dispersion curves for the five different equivalent depths of h = 8, 12, 25, 50, and 90 m. Source: Wheeler and Kiladis (1999) © American Meteorological Society. Used with permission.

Because EWs are coupled to convection, they influence the variability of rainfall (e.g., Wheeler and Kiladis, 1999; Straub and Kiladis, 2002; Mekonnen and Thorncroft, 2016; Schlueter et al., 019b; Ferrett et al., 2020; Ayesiga et al., 2021, 2022). The rate and time step of modulation is different for each wave. In terms of influencing convection variability, Wheeler and Kiladis (1999) ranks the MJO and Kelvin highest, followed by ER, MRG and IG. Over the Sahel, Schlueter et al. (019c) find that at time scales of 7–20 days, the MJO and ER waves are the most important while Kelvin waves are the dominate at 3-hourly to daily time scales explaining between 10–30% rainfall anomalies. They also found that the EIG and MRG waves were relatively unimportant. Over EEA, Mekonnen and Thorncroft (2016) showed that Kelvin waves influence the rainfall over western and eastern equatorial Africa. Their findings are confirmed by Ayesiga et al. (2021) who noted that Kelvin waves account for up to 5mm day^{-1} rainfall variation and that the effect is more frequent in the MAM rain season. Unfortunately, studies evaluating the ability of models to represent EWs paint a grim picture. Both Straub et al. (2010) and Huang et al. (2012) showed that 75% of models in Coupled Model Intercomparison Project phase 3 (CMIP3) did not adequately represent Kelvin waves. However, with improvement in models in recent times, the deficiencies in representing the EWs seem to be reducing as suggested by Yang et al. (2021). They found that Met Office global forecast model has substantial skill up to day 4 and day 6 for Kelvin and westward mixed Rossby–gravity (WMRG) waves, respectively. In conclusion, it is recommended to monitor the day to day activity of EWs to improve the accuracy of forecasts (e.g., Ayesiga et al., 2022).

2.4 Weather forecasting in EEA

2.4.1 Overview

Despite the improvement in weather forecasting in recent times, its benefits have not been realised in the tropics (Youds et al., 2021).This is despite the fact that Africa is one of the regions that urgently needs accurate weather information given her population is heavily reliant of rain-fed agriculture. IPCC (2022) notes that the continent is among the most vulnerable to extreme weather events. More optimistically, Parker et al. (2022) suggest that the continent is about to see an upturn in quality and usage of weather forecasts.

Weather forecasting is done for different forecast horizons, namely; nowcasting (usually up to 6 hours ahead), synoptic and short-range severe weather prediction (1–5 days), subseasonal-to-seasonal (2–6 weeks) and seasonal forecasting (months). All the forecasts in these horizons are important and should be undertaken together.
The following sections discuss the synoptic and short-range forecast horizon which is the horizon analysed in this work.

2.4.2 Synoptic and short-to-medium range forecasting

This forecast horizon caters for 1–14 days lead time. The forecasts are typically generated using numerical weather prediction (NWP) systems. Unfortunately over Africa, the skill of forecasts from these models is poor, hardly beating climatology (Vogel et al., 2018, 2020). Haiden et al. (2012) showed the gap between skill in the tropics and the extra-tropics is huge, with a 1 day forecast in the former being equivalent to a 6 day forecast in the latter. Over west Africa, it has been shown that a simple statistical forecasts beats state-of-the-art NWP forecasts (Vogel et al., 2021; Satheesh et al., 2023). To understand why the forecasts are bad over Africa, a brief background of NWP is explained in the next paragraph.

2.4.2.1 Historical perspective

As summarised in Parker and Diop-Kane (2017), NWP grew from three factors, namely; the historical behavior of observing the weather, theoretical principles in fluid dynamics by scientists like Isaac Newton, Vilhelm Bjerknes and many others, and development in computing resources. The first attempt to numerically produce weather forecasts was by Lewis Fry Richardson (Richardson, 1922). His attempt was largely deemed a failure. He did however demonstrated that the differential equations can be approximated as a set of algebraic equations to show the tendency of weather fields in space. These tendencies could then be extrapolated further in the future in small increments. His method was mechanical and required a force of 64000 people to cover the weather over the globe, signifying the need for a lot of computing power needed today by the state-of-art NWP systems. Following further developments in the field, the first NWP using a computer was produced by Charney, Fjortoft and von Neumann in 1950. Their model was based on the barotropic and baroclinic dynamics laid down by Charney, Rosby and Eady in 1940, 1947 and 1949, respectively. Since then, NWP has seen rapid development thanks largely to improvements in computing power and availability of observations.

2.4.2.2 NWP over Africa and the challenge of quality observational data

Over Africa, while the trajectory of development in NWP is similar to described above, the skill of the forecasts is still poor. Many explanations for this have been suggested. For example, Parker and Diop-Kane (2017) state that the mixed nature of weather regimes in Africa, comprising of mid-latitude weather systems, equatorial rain forests and deserts on different spatio-temporal scales make forecasting the weather a difficult problem. NWP models are known to struggle in representing convectively-triggered weather mainly because of poor parameterization (e.g., Marsham et al., 2013; Vogel et al., 2018; Birch et al., 2014). Over Africa, convective rainfall is the most common type. There is also the fact that the weather over Africa is highly variable, even at subseasonal time scales (e.g., de Andrade et al., 2021).

Because weather forecasting is an 'initial condition problem' (i.e., the observations and first guess are very important), quality observations are needed to initialise the model. Observations are also important for validation of the models which enables further model development (Tompkins et al., 2005; Faccani et al., 2009; Bormann et al., 2019). However, over Africa, quality and consistent observations are not available (Diem et al., 2014; Monsieurs et al., 2018). In fact, the station network coverage across Africa has been shrinking over the years (Asadullah et al., 2008; Dinku, 2019). Field campaigns over Africa have shown that a high density of observations and high resolution models improve the performance of NWP models (e.g., Faccani et al., 2009; Bormann et al., 2019). As shown in Fig. 2.9, the continent has the fewest number of stations and the lowest reporting frequency, implying that there is very little data to initialise the models over Africa. Moreover, in their experiments with ECMWF system, Bormann et al. (2019) showed that conventional observations, such as those from synoptic stations, have the highest influence on the performance of the model compared to other types of observations assimilated.

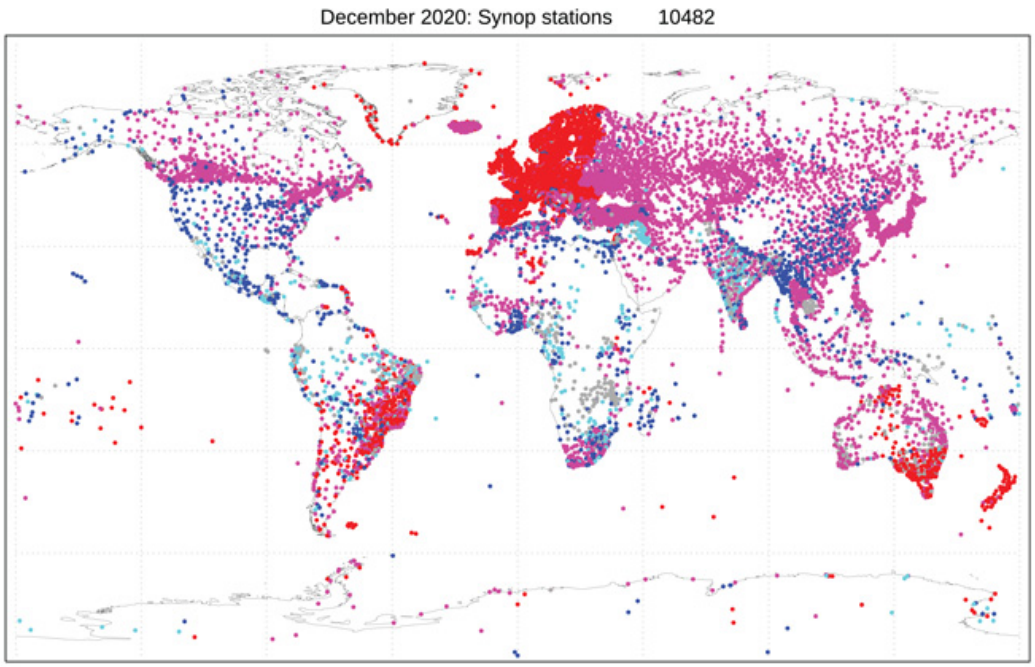


Figure 2.9: Active land surface stations (SYNOPs), December 2020. The colour gives the frequency of reports received: grey - fewer than 2/day through to red - 24/day. The WMO Integrated Global Observing System (WIGOS) is now encouraging the global exchange of hourly reports. Source: ECMWF.

To remedy the lack of quality observations over Africa, most forecasting centres assimilate satellite data. For rainfall, Satellite rainfall estimates (SREs) have filled the spatiotemporal data gaps. In most cases, the SREs are derived from radiances in the visible, infrared (IR), and microwave spectra measured by satellites. Singularly

or in combination, these retrievals provide high-resolution rainfall estimates, with a prime example being the Integrated Multi-satellite Retrieval for Global Precipitation Measurement (GPM) (IMERG) V06B dataset (Huffman et al., 2020). However, SREs are not without shortcomings. Despite having a wide field of view and high temporal availability, IR-based techniques rely on a cloud top temperature-precipitation relationship, which is often unsuitable for non-convective precipitation (Kidd and Huffman, 2011) or non-precipitating cold cirrus shields (Young et al., 2014). Microwave-based satellites, in turn, have been found to struggle with retrieving rain from warm clouds (Dinku et al., 010b; Monsieurs et al., 2018; Maranan et al., 2020) and to overestimate rainfall in places where convective rainfall dominates (Tian et al., 2009).

Validations of SREs over EEA have revealed that; a) SREs struggle retrieving rainfall over the complex topography of the region (e.g., Dinku et al., 010b; Diem et al., 2014; Monsieurs et al., 2018; Ageet et al., 2022), b) the skill is better for products that are calibrated with gauge data (e.g., Awange et al., 2015; Ageet et al., 2022), c) the rainfall formation process also influences skill (Mccollum et al., 2000; Maranan et al., 2020). Therefore, the SREs product used for validation of forecasts must be chosen carefully, hence, the need to validate these products prior to their use.
Overall, to improve the synoptic and short-to medium range forecasts from NWP systems, more observations have to be included. The better observation network over Europe and America (Fig.2.9) partly explains the better skill of NWP forecasts over the extratropics. With the introduction of Systematic Observations Financing Facility (SOFF), WMO is trying to improve the global observation network, especially inplaces such as Africa WMO (2023). Besides improving the observation network, Parker et al. (2022) also recommend using convection-permitting models and ensembles to improve rainfall forecasts from NWP among other solutions.

3. Research questions

One of the fundamental shortcomings in the region that contributes to poor forecasts is the low number and low quality of ground observation data sets (e.g., Maidment et al., 2013; Diem et al., 2014). As shown by Nicholson et al. (018b) and and Dinku (2019), the network of gauges has continuously shrunk over the years. Even where observations are available, the data rarely makes it to the global telecommunication system (GTS) on time to be assimilated into weather forecast models or included in the gauge calibration of satellite rainfall estimates. Even model development is hindered due to lack of quality data to do research. The next available option in recent times is remotely sensed observations, especially in the form of SREs. However, as discussed in the last two chapters, despite having an excellent spatio-temporal coverage, these estimates have their shortcomings too. For example, SREs often retrieve rainfall in places covered with non-precipitating cirrus clouds, especially in the IR channel (e.g., Young et al., 2014; Maranan et al., 2020). The microwave channel has been found to struggle with retrieving warm rain (e.g., Monsieurs et al., 2018; Maranan et al., 2020) while overestimating rainfall in regions where convective rainfall dominates, e.g., Lake Victoria (Tian et al., 2009). Therefore, there is need to ensure that the most suitable SRE is used for any research. As a result, the first research aim was to identify the best performing SREs in the study region. This aim was addressed by answering the following set of questions:

RQ 1a What is the best performing SRE over EEA at daily, pentadal, decadal, monthly and annual time scales?

RQ 1b Does the performance of SREs vary spatially over EEA?

RQ 1c Are the SREs capable of retrieving extreme rainfall events over EEA and, if so, which is the best product?

Having identified the most suitable SRE to use going forward, the next aim was to assess the quality of global weather forecast products over EEA. Previous research shows that skill is poor in the tropics (e.g., Haiden et al., 2012; Vogel et al., 2018, 2020). However, the performance of forecasts is not homogeneous. Some studies have found that the

skill, particularly over EEA is better compared to other regions of Africa (de Andrade et al., 2021). This good skill is attributed to advances in computing and hence, more data and complex models can be used. Recently, more sources of predictability have been discovered, e.g., the MJO (Vitart et al., 2017; de Andrade et al., 2021). These gains too are not homogeneous across the region and forecast products. Most studies showed that the ECMWF forecast is the best-performing product over EEA, hence the choice to use this model. Specifically, reforecasts were used, which have the benefit of using the latest model-data assimilation system and a constant forecast system over time. Given the questionable skill and global nature of the forecast products, they are prone to errors in the form of miscalibration and generalisation (e.g., Vogel et al., 2020). A common practice in weather forecasting is to postprocess the raw forecast to correct for systematic biases. This process should be carried out by all National Meteorological and Hydrological Services (NMHS) to ensure that the forecasts they issue are tuned to the local conditions. The NMHS are slowly getting expertise to do this task and, with computing resources becoming available and novel simple methods of postprocessing like isotonic distributional regression (IDR; Henzi et al., 2021), we should expect better forecasts in the near future. To achieve the second aim, the following set of questions is asked:

RQ 2a Does the ECMWF deterministic reforecast correctly estimate the intensity of rainfall over EEA?

RQ 2b Do the raw ensemble ECMWF reforecasts predict the correct occurrence of a rainy day and rainfall amounts and are they better than a reference forecast?

RQ 2c Do the reforecasts have any systematic biases and can postprocessing using IDR alleviate the inadequacy?

RQ 2d Are there differences in the performance of the reforecasts over the different topographical features of EEA?

Having ascertained the skillfulness or lack of skill of the reforecasts, it is instructive to determine the sources of the observed performance. Weather has a predictability limit which varies spatio-temporally. For example, although short-term local forecasts are poor in the tropics, we know that the predictability limit is longer there compared to the extratropics (Judt, 2020). This is due to existence of propagating signals, for example EWs and MJO. Indeed, at the subseasonal and seasonal scales, the forecasts in the tropics are relatively better compared to the short-to-medium range forecasts. From the previous chapter, it is clear that the weather and climate of EEA is variable (e.g., Phillips and Mcintyre, 2000; Nicholson, 2017; Seregina et al., 2018). This variability is controlled by the different weather and climate drivers previously explained. Although

these drivers upset the normal pattern of the weather and climate, they are sources of predictability. For example, the improved forecasts of the MJO in global models has been credited for the leap in performance of forecasts over Africa. Since modes like ENSO and IOD develop over months, they are usually accurately forecast, offering an opportunity for improved forecasts. This therefore motivates the third research aim, which is to assess whether the different weather regimes play a role in the skill of weather forecasts. To achieve this aim, the following questions were asked:

RQ 3a Do EWs and the MJO modulate rainfall over EEA and by how much?

RQ 3b Do EWs and the MJO influence the skill of the ECMWF reforecasts over EEA?

RQ 3c Does the IOD modulate rainfall over EEA? What is their influence on the skill of ECMWF reforecasts?

4. Quality of satellite rainfall estimates over Equatorial East Africa

This chapter presents finding of the assessment of SREs against a relatively dense and quality-checked gauge observation network. First, the study domain, datasets and methods are explained in section 4.1. Then the results of the validation at multiple spatio-temporal scales, i.e., regional and whole-domain, and annual, monthly, dekadal, pentadal and daily aggregations are presented in section 4.2. The results of the analysis of the SREs' ability to retrieve extreme rainfall events, and hence, their feasibility for estimating return periods of the extremes are also presented in this section. This chapter ends with section 4.3 where the discussion and conclusion of the results herein were put in broad perspective and compared with previous studies.

4.1 Study area, datasets and methods description

4.1.1 Study domain

The study focuses on Uganda and the surrounding areas (Fig. 4.1). The topography is heterogeneous with elevation ranging from 375m to 5109m above mean sea level at Lake Turkana in the Northeast and the peak of Mt. Rwenzori in the West, respectively. It is also composed of a number of large inland lakes, mountains, parts of the Great East African Rift Valley, and rivers. Most parts of the region exhibit two distinct rainfall seasons (March–May and September–November), locally referred to as the 'long rain season' and 'short rain season' with peaks in April and October, respectively (Nicholson, 2017; Seregina et al., 2018). Over Western Uganda, a transitional region between the eastern and western equatorial Africa, the MAM and SON seasons form the short and long rains, respectively (Diem et al., 2019). The northern part of the study area exhibits a unimodal rainfall peak in May through August (Phillips and Mcintyre, 2000; Seregina et al., 2018). As discussed in Nicholson (2017), the climate of the region is controlled by three classes of factors; regional forcing features (e.g., Walker circulation, El Niño-Southern Oscillation, Indian Ocean Dipole and Madden-Julian Oscillation), regional circulation features (e.g., tropical easterly jet, low level westerlies and monsoons), and local geographic factors (e.g., Lake Victoria, East African highlands). These factors

inevitably lead to variability in rainfall and coupled with the heterogenous topography present a challenge to satellite rainfall retrievals (e.g., Monsieurs et al., 2018; Coz and van de Giesen, 2020).

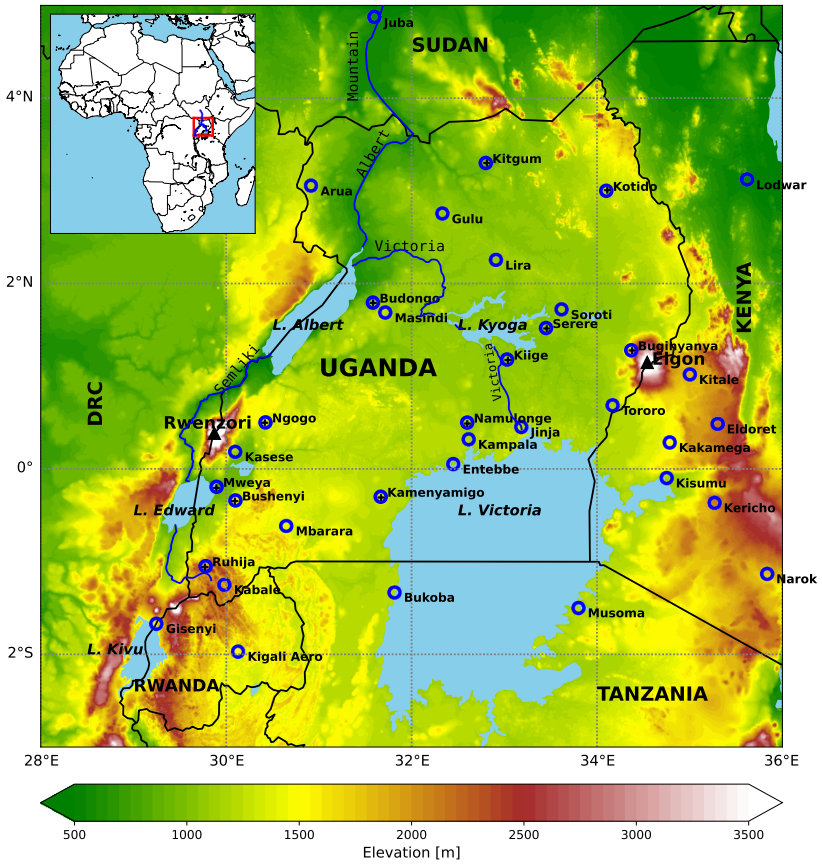


Figure 4.1: The study area with location and names of the rain-gauge stations used in the study shown with blue circles. Data from stations with black cross in the centre are not transmitted to the Global Telecommunication System (GTS). The major lakes, rivers, and mountains, viz. Rwenzori in the west and Elgon in the east are also indicated. The color shading shows the elevation from Global Land One-km Base Elevation Project (GLOBE; Hartigan and Coauthors, 1999). The map inset highlights the location of Uganda within Africa (red box). Taken from Ageet et al. (2022). © American Meteorological Society. Used with permission.

4.1.2 Datasets

4.1.2.1 Satellite estimates

We included eight satellite datasets in this validation study (Table 4.1). The temporal resolutions of the datasets vary from 30-minute to daily, while the spatial resolutions vary from 0.1°to 0.25°latitude-longitude grid boxes. They are described in the following in more detail.

TMPA V07 (Huffman et al., 2007, 2018) is a multi-satellite precipitation product with latencies of about 8 hours and about 2 months for the 'real time' and 'final'

Table 4.1: Summary of datasets used in the study. The period and re-gridded resolution used in this study is given in parentheses in the 'temporal coverage' and native resolution' columns, respectively. Taken from Ageet et al. (2022). © American Meteorological Society. Used with permission.

Dataset	Full name	Native spacial resolution	Temporal resolution		Latenc	Temporal coverage
Reference						
TMPA 34B42 v7	Tropical Rainfall Measuring Mission (TRMM) Multisatellite precipitation Analysis	0.25°	3-hourly	2 months	January 1998–present (2001–18)	Huffman et al. (2007), Huffman et al. (2018)
TMPA-RT 3B42RT v7	47	877	230			
IMERG	Integrated Multi-satellitE Retrieval for GPM (Global Precipitation Measurement)	0.1°(0.25°)	Half-hourly	3.5 months	June 2000–Present (2001-2018)	Huffman et al. (020b)
IMERG-E	Integrated Multi-satellitE Retrieval for GPM (Global Precipitation Measurement)	0.1°(0.25°)	Half-hourly	4 hours	June 2000–Present (2001-2018)	Huffman et al. (020c)
CHIRPS	Climate Hazard Infrared Precipitation with Stations	0.05°/0.25°	3-hourly	2 months	January 1981–Present (2001-2018)	Funk et al. (2015)
MSWEP v2.2	Multi-Source Weighted-Ensemble Precipitation V2.2	0.1°(0.25°)	3-hourly	Irregular	January 1979–October 2017 (2001-2016)	Beck et al. (2019)
MSWEPv2.8	Multi-Source Weighted-Ensemble Precipitation V2.8	0.1°(0.25°)	3-hourly	Irregular	January 1979-December 2020 (2001-2018)	Beck et al. (2019)

runs, respectively. TMPA consists of PMW precipitation estimates, IR precipitation estimates, and GPCP monthly precipitation. First, PMW estimates are calculated using the Goddard Profiling (GPROF) Algorithm (Kummerow et al., 1996, 2001). Then all available PMW estimates are combined and completed with MW-calibrated IR data in case of existing gaps. Finally, monthly multi-satellite (MS) accumulations are combined with Global Precipitation Climatology Centre (GPCC) monthly gauge analysis to create a satellite-gauge (SG) monthly product (TMPA 3B43). A SG/MS ratio is then calculated for each 0.25°and used to calibrate each 3-hourly field of the month, producing the final 3-hourly product (TMPA 3B42).

IMERG V06B (Huffman et al., 2020) is a MS precipitation product with latencies of about 4 hours, 14 hours, and 3.5 months for the 'early', 'late' and 'final' runs, respectively. IMERG builds on the TRMM legacy and uses similar data inputs as TMPA, with the GPM core observatory replacing the TRMM counterpart. The PMW estimates are calculated using a more recent version of the GPROF algorithm. A seasonal GPCP calibration is applied to the PMW estimates to yield 30-minute 0.1°×0.1°fields. These fields are then spatiotemporally morphed to fill the gaps for areas without PMW overpasses. This step results in the MS, half-hourly IMERG early and IMERG late products ('IMERG-E' and 'IMERG-L' henceforth). IMERG-E contains only forward morphing while IMERG-L has both forward and backward morphing. Monthly satellite-gauge estimates are created by summing the half-hourly estimates for the month and calibrating with GPCP monthly precipitation analysis. Finally, gauge calibration of the half-hourly estimates is exercised in a similar fashion to TMPA to create the final version of IMERG ('IMERG-F' hereinafter).

CHIRPS v2 ('CHIRPS' hereinafter) is a quasi-global dataset providing daily, pentadal and monthly estimates (Funk et al., 2015). CHIRPS is based on a global monthly precipitation climatology (CHPclim), thermal infrared (TIR) Cold Cloud Duration (CCD), and daily and monthly RG data. TIR CCD data created at a constant CCD temperature threshold of 235, is locally calibrated using TMPA 3B42 pentadal precipitation. The pentadal estimates are then multiplied by their corresponding CHPclim estimate to produce Climate Hazard Infrared Precipitation (CHIRP). CHIRP is merged with gauge data to produce CHIRPS on pentadal and monthly timesteps. The pentadal CHIRPS is disaggregated into daily CHIRPS precipitation using daily CCD data.

MSWEP is a global precipitation dataset available from 1979 to October 2017 (MSWEPv2.2, Beck et al., 2019) and until December 2020 in a recent update (MSWEPv2.8, Beck et al., 021a). The dataset is created through weighted merging of gauge-, satellite- and reanalysis-based precipitation products. The merging weights are determined by assessing the performance of the input datasets individually against quality-controlled gauge precipitation. The weighted merging results into a 3-hourly reference precipitation dataset which is then calibrated using daily and monthly gauge precipitation to create

MSWEPv2.2. Although the algorithm did not change substantially, changes in the production of MSWEPv2.8 include a reduction to two underlying datasets from the five used in MSWEPv2.2 and a reduction in the amount of gauge data used (Beck et al., 021b).

4.1.2.2 Rain gauge rainfall

Besides quality issues, not nearly enough data is ingested into the GTS in our study area. For this study, 24 of the 36 stations transmitted rainfall data to the GTS, and only 6 of these 24 stations had more than 50% of available daily data for the study period (Fig. 4.2a). Benefitting from collaborations, the Karlsruhe African Surface Station Database (KASS-D; Vogel et al., 2018) has many observations not reported to the GTS. Figure 4.2b shows the full station ensemble and its data availability in KASS-D within the study region. Hence, 12 non-GTS stations are available for the present analysis and even the GTS reporting stations have much more complete data in KASS-D.

The quality control of the gauge dataset was carried out following two steps. First, the timestamp was checked using a methodology similar to Beck et al. (2019) to ensure a consistent treatment of the timestamp across all datasets. The datasets were then analysed for suspicious outliers and zero values using a combination of techniques, e.g., using the 'letter-value plot' method in the python package "seaborn" and accumulation curves.

4.1.3 Methods

4.1.3.1 Point-to-pixel comparison

A point-to-pixel approach is performed by comparing point gauge data to the closest satellite pixel value (e.g., Monsieurs et al., 2018; Maranan et al., 2020). We recognize the shortcomings of this approach given that gauges are highly localised and may not be representative of a grid-box coverage. Engel et al. (2017) and Monsieurs et al. (2018) demonstrate the shortcoming using grids with multiple RGs. We have not carried out a similar analysis given our dataset has no grid boxes with more than one RG. We instead retain the IMERG native resolution (0.1°) and compare it with the remapped 0.25°resolution to assess changes in the performance of SREs due to changes in resolution. While interpolation of irregular station data to the SRE grids is recommended for satellite validation studies (Maidment et al., 2013), we did not use this approach due to the relatively low density of rain gauges and the complex topography. All the SREs are re-gridded to the 0.25°resolution of TMPA using a first order conservative remapping (Jones, 1999) to allow for a fair comparison of the SREs. Since gauge data is missing at some stations for some days, we only used satellite data when gauge data were available at a given station and day.

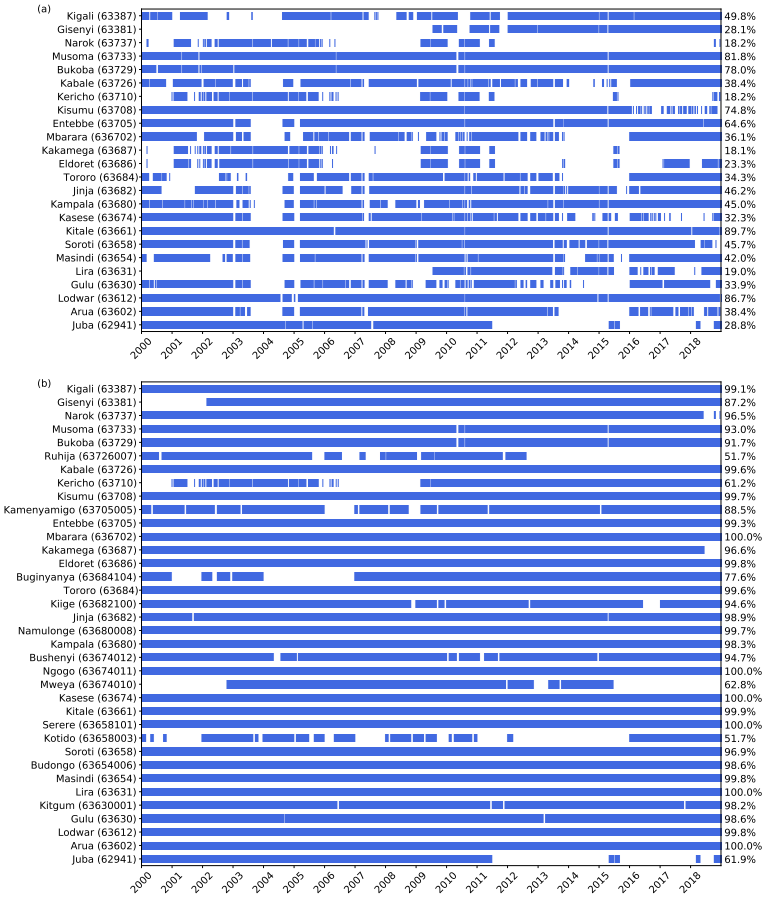


Figure 4.2: Temporal availability of the gauge data used in the study with stations reporting to the GTS (a) and the full dataset as obtained from KASS-D (b) for the period 2000-2018. The GTS-stations subset has only 24 stations, identifiable by their five-digit WMO station numbers while the full dataset obtained from KASS-D has 36. The blue-strips shading show daily data availability with the percentage for each station during the period 2000–2018 indicated by the numbers on the right. Taken from Ageet et al. (2022). © American Meteorological Society. Used with permission.

4.1.3.2 Validation metrics

This study employs some commonly used standard validation approaches (e.g., Ebert, 2007; Wilks, 2011). First, Quantile-Quantile (QQ) plots were used to compare the distribution of rain rates. The ability of SREs to detect rainy days, defined here as days with rainfall total of more than 0.2 mm, was assessed based on the contingency table (Table 4.2) using dichotomous metrics, namely, probability of detection (POD), probability of false alarm (POFA), bias in detection (BID), and Heidke skill score (HSS). For analysis of extremes, we applied a threshold (95th percentile of the subset of rainy days in the RGs) to differentiate between an “extreme” and “non-extreme” rainfall day. Thus, a hit occurs when both the RG and SRE have a rainfall amount greater than the threshold. The second group of metrics includes the Pearson’s correlation coefficient (r), mean error (ME), percent bias (PB), mean absolute error (MAE) and root-mean

square error (RMSE) to assess the SREs' accuracy for rain rates using the subset of hits. Additionally, the Nash-Sutcliffe coefficient of efficiency, E (Nash and Sutcliffe, 1970; Legates and Mccabe, 1999) was used to assess the skill against climatology. The metrics are summarised in Table 4.3.

Table 4.2: 2 x 2 contingency table for comparing the rainy days in gauge and satellite estimate. A day is considered as rainy if 0.2 mm of rainfall or more is recorded. Note that for the when a threshold is set, a wet day occur when an amount greater than the set threshold was recorded. Taken from Ageet et al. (2022). © American Meteorological Society. Used with permission.

	Gauge $\geq 0.2mm$	Gauge $< 0.2mm$
Satellite $\geq 0.2mm$	Hit (H)	False alarm (F)
Satellite $< 0.2mm$	Miss (M)	Correct rejection (R)

4.1.3.3 Spatio-temporal analysis

The validation was performed for individual stations to fully use the available gauge data and for daily, pentadal, dekadal, monthly and annual aggregations. To reduce randomness in some of the analyses, we reorganized the stations into groups. We therefore used the non-hierarchical K-means clustering algorithm (Pedregosa et al., 2012) to partition the stations into groups. This approach aims at placing the stations in clusters such that the intra-cluster variance is minimized (Hartigan and Wong, 1979). As noted by Cattani et al. (2016), K-means clustering is prone to subjectivity since the maximum number of clusters has to be determined a priori. We used the 'elbow' method (Thorndike, 1953; Zhang et al., 2016) which is based on the intra-cluster sum of squared differences to determine the optimum number of clusters. For each station, a 30-day running mean was applied over their respective Julian-day-of-year-based climatology of rainfall. The resulting time series were then normalized with the maximum value of the 30-day running mean at the respective stations to ensure that stations with similar seasonal cycles are grouped together regardless of the amount of precipitation received. For each cluster, the mean monthly rainfall of the satellite products and RGs were computed by taking the average of the stations in that cluster. Additionally, to investigate the influence of gauge calibration, we subset the data into two groups, namely, GTS (stations with the highest percentage of data reported to the GTS) and non-GTS (stations whose data is not reported to the GTS). In order to allow for a comparison of stations subsets with different climatologies, we normalized the error metrics (ME, MAE and RMSE) with the standard deviation of the respective RGs before taking the mean over all stations in the subset.

Table 4.3: Summary of the validation metrics used. H, F, M and R denote hits, false alarms, misses and correct rejections, respectively. The x_i, y_i,$\bar{x}$ and $\bar{y}$ terms represent the rainfall totals at the gauge for a given time, SRE rainfall totals for a given time, the mean rainfall totals at gauge, and the mean rainfall totals of SRE, respectively. The range and best possible score of each statistic is added in the first column in square brackets and bold numbers, respectively.

Statistic	Formula		Description
POD, [0,1], **1**	$POD = \frac{H}{H+M}$	(4.1)	Ability of the SREs to correctly identify rainy days
POFA, [0,1], **0**	$POFA = \frac{F}{H+F}$	(4.2)	Proportion of rainy days in the SREs that were not observed in RGs
BID, [-∞, +∞], **1**	$BID = \frac{H+F}{H+M}$	(4.3)	Assesses whether the SRE overestimates or underestimates rainy day frequency
HSS, [-∞, 1], **1**	$HSS = \frac{2(HR-FM)}{(H+M)(M+R)+(H+F)(F+R)}$	(4.4)	Assesses the skill of SREs compared to random chance
r, [-1,1], **1**	$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})\sum_{i=1}^{n}(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$	(4.5)	Assesses the covariance of the gauge data with that of the SRE
ME (mm), [-∞, ∞], **0**	$ME = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)$	(4.6)	Measures the bias and its direction (underestimation or overestimation) by the SREs
PB (%), [-∞, ∞], **0**	$PB = 100\frac{\frac{1}{n}\sum_{i=1}^{n}(y_i-x_i)}{\bar{x}}$	(4.7)	Assesses the tendency of the SREs to overestimate or underestimate rain rates relative to the mean of the RGs
MAE (mm), [0,∞], **0**	$MAE = \frac{1}{n}\sum_{i=1}^{n}\lvert(y_i - x_i)\rvert$	(4.8)	Measures the error of the SREs regardless of direction
RMSE (mm), [0,∞], **0**	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2}$	(4.9)	Measures the error of the SREs but assigning more weight to outliers
E, [-∞, 1], **1**	$E = 1 - \frac{\sum_{i=1}^{n}(x_i-y_i)^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$	(4.10)	Assesses the skill of the SREs relative to climatology

4.1.3.4 Extreme rainfall events analysis

Since hydrometeorological hazards are largely a result of extreme daily rainfall events, the ability of satellites to capture these events is analysed. We define an "extreme event" as daily rainfall totals exceeding the 95th percentile on rainy days in the RG dataset. The extreme events thresholds were computed at individual stations and the corresponding values for the SREs were obtained from the closest pixels. We also analysed SREs' suitability in simulating the return values of extreme event. Extreme value models are used to study the behaviour of the tail of a distribution (Bommier, 2014), enabling the simulation of extreme values and their return periods. Similar to Engel et al. (2017), we use the peak-over-threshold (POT) method to fit the generalised Pareto distribution (GPD; Lemos et al., 2020) to the RG and SREs daily rainfall. The GPD was fitted to the subsets of extreme events (i.e., >95th percentile) in the RG and SREs data sets. The extremes in the SREs were obtained in a similar way to the RGs extremes. To make the stations and all the rainfall products comparable, we normalized the modelled return values with the RG-modelled return values at the stations, then averaged over all stations for each data set. The normalized return values of the RG data were taken as the reference for evaluating the SREs. The performance of SREs was considered 'satisfactory' if the return value modelled using the SRE fall within the standard deviation of the RG's return value.

4.2 Results

While we present results of all the eight SREs products on the same figures and tables, the comparison between the early products (IMERG-E and TMPA-RT) is discussed separately, as is the comparison between the early and the respective final versions (IMERG-F and TMPA). Furthermore, a separate analysis of the influence of different spatial resolutions on the skill is conducted using IMERG-F at 0.25°and at its native resolution 0.1°.

4.2.1 Annual scale

Based on RGs, the study area receives an annual average of 1217 mm with high spatial variability. The wettest region is the eastern part and shorelines of Lake Victoria (Fig. 4.3). The Ugandan Ssese Islands are known to have more than 2000 mm yr^{-1} (Flohn and Fraedrich, 1966) with a short-term measurement on the Tanzanian Nabuyongo Island in the middle of the lake indicating the possibility of 3000 mm yr^{-1} (Flohn and Burkhardt, 1985). No rainfall data over the lake are available in the study period, only Bukoba at the eastern shore with 1924 mm yr^{-1} indicates the wetness of this area. Other wet regions are the Rwenzori and Elgon Mountains in western and eastern Uganda, respectively. While no station is available in the Rwenzori mountain ranges for this study, amounts of 2000–3000 mm yr^{-1} are reported (Eggermont et al., 2009). Buginyanya at Mt. Elgon

at an elevation of 1845 meters receives an annual total of 2278 mm. From Mt. Elgon, a wetter region with annual totals of more than 1200mm stretches northwestward across the Lake Kyoga region in central Uganda to northwestern Uganda west of the Albert Nile (Fig. 4.3; Basalirwa, 1995). The northeastern and southwestern parts of the study area are located along the East African Rift Valley, an area locally known as the 'cattle corridor', which are generally associated with drier semi-arid conditions. Here, rain shadow effects are the likely causes of the low rainfall totals in low-elevation areas (e.g., Diem et al., 2014).

The satellite products generally reproduce the spatial pattern of annual rainfall, with CHIRPS showing the best performance for most of the scores (Fig. 4.3 and Table 4.4). The correlations between the RGs and satellites' mean annual rainfall totals are high, with CHIRPS and MSWEPv2.2 having the highest correlation of 0.84, followed by MSWEPv2.8 (0.82), IMERG-F (0.78), and then TMPA (0.75, Table 4.4). All the satellite products are skilful in depicting annual rainfall totals with all of them having an E value greater than 0.5. Although the patterns are relatively well reproduced, biases do exist. For example, when averaged over all stations, IMERG-F overestimate the annual rainfall by about 5%, while TMPA, CHIRPS, MSWEPv2.2 and MSWEPv2.8 all underestimate annual rainfall by about 5%, 1%, 8% and 7%, respectively.

Table 4.4: Scores for annual rainfall totals (2001-2018, hence, total number of observations, N = 648) for all the satellite products, except for MSWEPv2.2 (2001-2016, hence, total number of observations, N = 576). The best performing final product for each metric is given in bold font. Taken from Ageet et al. (2022). © American Meteorological Society. Used with permission.

Statistic/Product	IMERG-F$_0$.10ř	IMERG-F-0.25°	IMERG-E	TMPA	TMPA-RT	CHIRPS	MSWEPv2.2	MSWEPv2.8
r	0.77	0.78	0.75	0.75	0.74	**0.84**	**0.84**	0.82
E	0.57	0.59	-0.36	0.52	0.23	**0.70**	0.65	0.63
ME (mm yr^{-1})	+57	+57	+349	-60	-214	**-10**	-99	-91
PB (%)	5	5	28	-5	-19	**-1**	-8	-7
MAE (mm yr^{-1})	208	200	402	218	300	171	**169**	193
RMSE (mm yr^{-1})	294	287	520	309	392	**243**	264	272

Two salient features in Fig. 4.3 are worthy of mentioning; the very high degree of wetness of the IMERG products over Lake Victoria with values over 3600 mm yr^{-1} (Figs. 4.3a,b), corroborating the findings of Nicholson et al. (2021), and the sharp horizontal discontinuity between the northern and southern parts of the lake in TMPA products (Figs. 4.3d,e). While IMERG remedies the discontinuity likely with an improved land-sea mask for the gauge calibration (see Fig. A.1), the overestimation with respect to RGs and the other SREs might be related to a larger weight given to PMW retrievals in IMERG that are known to overestimate rainfall in areas with deep convection (O and Kirstetter, 2018; Nicholson et al., 2021). IMERG-E and TMPA-RT also show quite high correlation values of 0.75 and 0.74, respectively, although the E value of IMERG-E suggests a worse performance than climatology. Additionally, IMERG-E

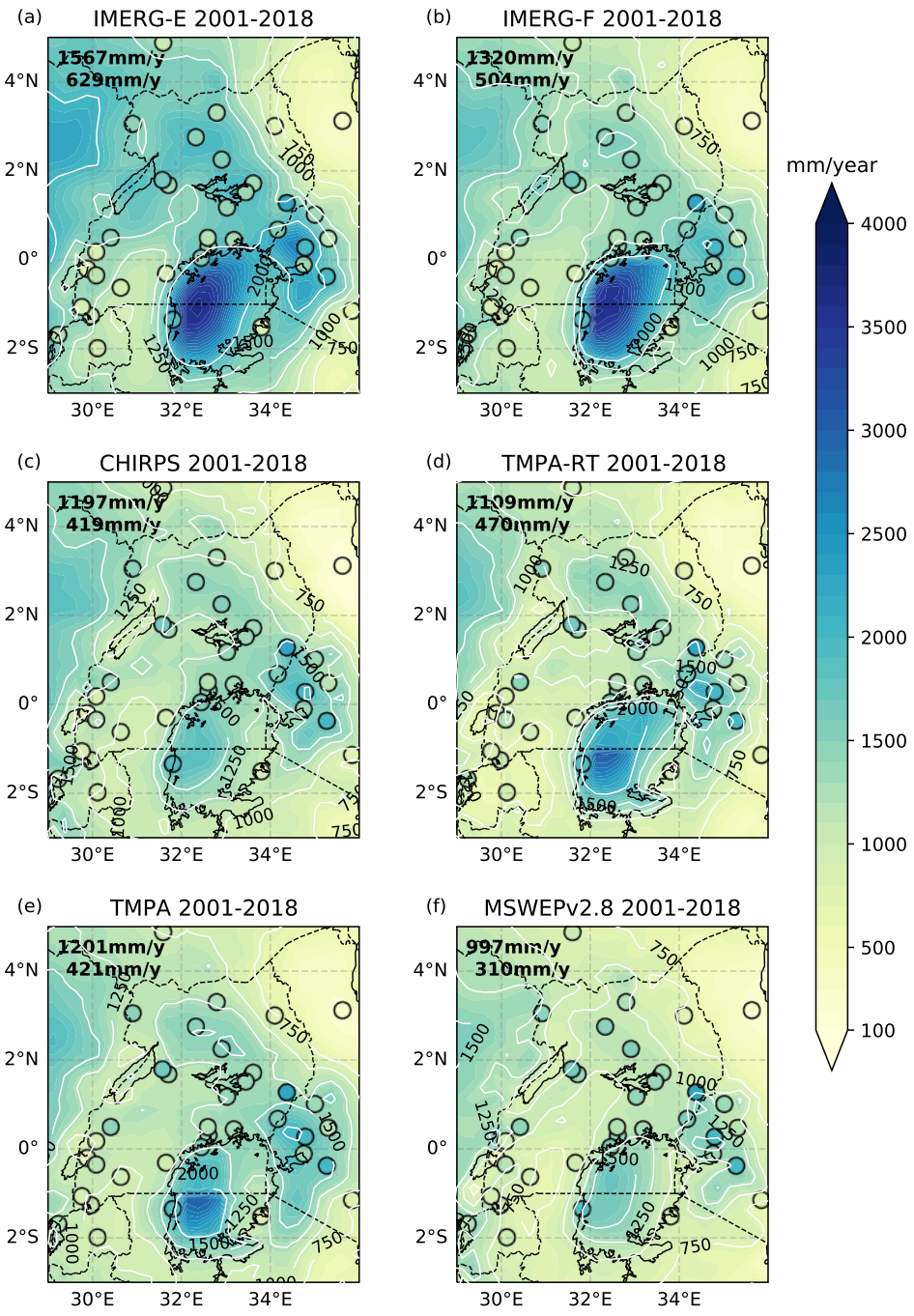


Figure 4.3: Mean annual rainfall totals for six satellite products (shaded) with mean annual rainfall totals of the RGs annotated in the circles. The white contours are isohyets of mean annual rainfall, labelled in black. The dashed black lines are country boundaries while the solid black lines demarcate inland lakes. The values on the top-left corners are the mean annual rainfall and the standard deviation taken over the entire domain for the respective satellite product. All the products are re-gridded to a 0.25°common grid. Taken from Ageet et al. (2022). © American Meteorological Society. Used with permission.

overestimates annual rainfall by 28% (349 mm yr^{-1}), while on the other hand TMPA-RT underestimates annual rainfall by 19% (214 mm yr^{-1}). The large difference between IMERG-E and TMPA-RT is likely due to a very significant difference in the algorithm between IMERG and TMPA. These larger errors in the early products compared to their respective final products underscore the importance of gauge calibration. Based on the domain-averaged annual rainfall of the SREs , MSWEPv2.8 retrieves the least annual rainfall of 997 mm with a standard deviation of 310 mm. The mean annual rainfall is substantially less than that retrieved by MSWEPv2.2 (1125 mm). Moreover, the r and E values show a decline in performance from MSWEPv2.2 to MSWEPv2.8 (Table 4.4).

4.2.2 Monthly time scale

Based on the elbow method (cf. Section 2c), the optimum number of clusters to be used in the K-means clustering was found to be four. The names of the clusters given in Fig. 4.4 were assigned based on the relative location of the majority of the stations forming the clusters (i.e., 'Northern' (N-) cluster, 'Central' (C-) cluster, 'Southwestern' (SW-) cluster, and 'Lake Victoria Basin' (LVB-) cluster). The four clusters depict different seasonal cycles (Figs. 4b–e). The N-cluster depicts a unimodal rainfall regime with the starting from April to October with two small peaks in May and August. These peaks are partly associated with the migration of the tropical rain belt during its meridional transitions (Seregina et al. 2019). The remaining three clusters have bimodal rainfall regimes. In the C-clusters the wet seasons run from March to June and August to November, with peaks in April and October. In the SW- and LVB-cluster, the first wet season occur from March to May, peaking in April. The second rainy seasons are September–November and October–December, for the SW- and LVB-cluster, respectively. The high overestimation of IMERG-E is clearly seen in all the clusters.

Generally, all the satellite products replicate the seasonal cycle well (Figs. 4.4b–e, A.2 and Table 4.5). The agreement between the RGs and the satellite products is evident from the high correlation values of between 0.77 (TMPA) and 0.81 (MSWEPv2.2). All the satellite products perform better than the reference RG-based climatology (i.e., $E > 0$; 0.58–0.65). Considering the final products only, IMERG-F overestimates monthly rainfall by about 4%, while MSWEPv2.2, MSWEPv2.8, TMPA and CHIRPS tend to underestimate rainfall by about 8%, 8%, 5% and 1%, respectively. The results for this time scale also point to a decline in performance of MSWEPv2.8 compared to MSWEPv2.2. The NRT products also reproduce the seasonality of rainfall in the study areas, but the errors are larger compared to their respective final versions. IMERG-E (TMPA-RT) overestimates (underestimates) monthly rainfall by about 28% (20%). Overall, MSWEPv2.2 outperforms all the other products at this timescale.

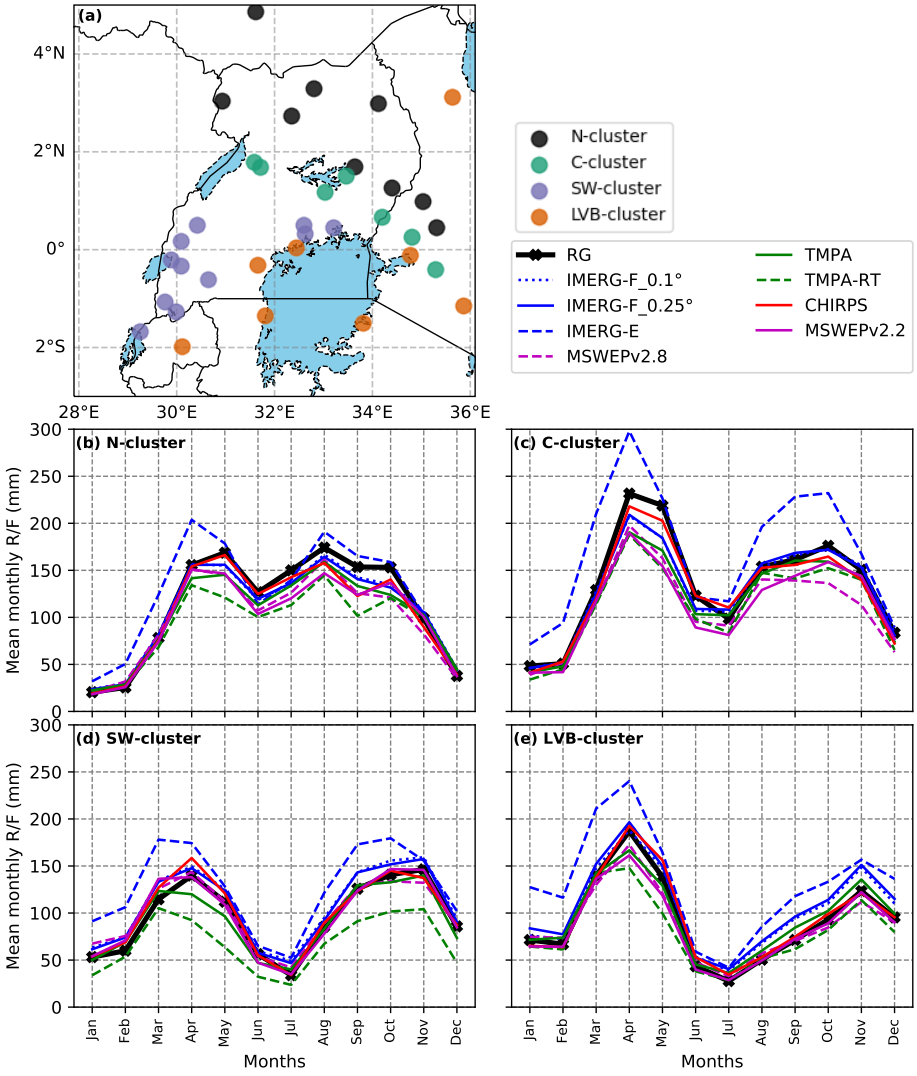


Figure 4.4: Result of the K-means clustering (a). Black-, green-, magenta- and orange- coloured stations indicate the 'Northern' cluster (N-cluster), 'Central' cluster (C-cluster), 'Southwestern' cluster (SW-cluster) and 'Lake Victoria Basin' cluster (LVB-cluster), respectively. Some stations may be relatively far from their cluster, e.g., the three SW cluster stations located at the shores of Lake Victoria. Manual investigation showed the seasonal cycles of these stations match the SW cluster. Panels (b)–(e) show the mean seasonal cycles of rain gauges and satellite products in the four clusters for the period 2001–2018 (IMERG-F-0.1°, IMERG-F-0.25°, IMERG-E, TMPA, TMPA-RT, CHIRPS and MSWEPv2.8) and 2001–2016 (MSWEPv2.2). Taken from Ageet et al. (2022). © American Meteorological Society. Used with permission.

4.2.3 Daily, pentadal and dekadal time scales

All the SREs, except MSWEPv2.2 slightly overestimate the low rain rates whereas high rainfall rates are largely underestimated. This is evident from the QQ plots in Fig. 4.5. MSWEPv2.8 visibly underestimates the higher rainfall rates compared to MSWEP v2.2 but overestimates the low rain rates. Although MSWEPv2.2 is below the diagonal

Table 4.5: Similar to Table 4.4, but for monthly rainfall (N = 7,776 for all SREs, except MSWEPv2.2 with N = 6,912). Taken from Ageet et al. (2022). © American Meteorological Society. Used with permission.

Statistic/Product	IMERG-F$_0$.10ř	IMERG-F-0.25°	IMERG-E	TMPA	TMPA-RT	CHIRPS	MSWEPv2.2	MSWEPv2.8
r	0.79	0.80	0.74	0.74	0.77	0.79	**0.81**	0.79
E	0.63	0.64	24	0.58	0.45	0.62	**0.65**	0.62
ME (mm yr^{-1})	+4.8	+4.8	+30.3	-5.3	-20.4	**-1.1**	-8.8	-8.1
PB (%)	4	4	28	-5	-20	**-1**	-8	-8
MAE (mm yr^{-1})	34.8	34.4	50.2	36.2	42.0	35.1	**31.0**	34.9
RMSE (mm yr^{-1})	49.2	48.6	70.1	52.0	59.6	49.0	**47.3**	49.6

for almost all the percentiles, it has a better fit to the RG data. The distribution in MSWEPv2.8 is likely strongly influenced by ERA5 (see Fig. A.3), given that their patterns are similar.

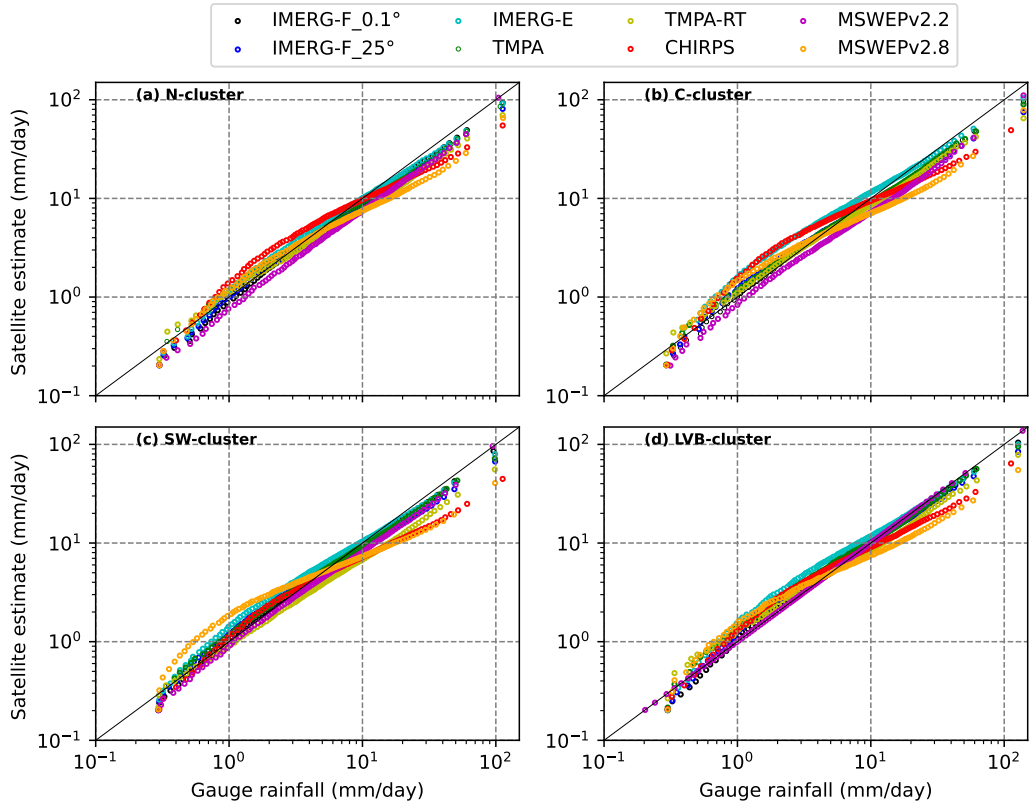


Figure 4.5: QQ plot of daily gauge rain rates against daily satellite rain rates in the four clusters. The solid diagonal line is the 1:1 line (ideal fit). Note that the scales are logarithmic and only hits are considered (days when both SRE and RG > 0.2 mm). Taken from Ageet et al. (2022). © American Meteorological Society. Used with permission.

The ability of the final versions of the satellite products to correctly capture a rainy day as reported in the gauges is generally good. This is illustrated by high POD values ranging from 0.76 in TMPA to 0.97 in MSWEPv2.8 (Table 4.6). However, all the products are prone to false alarms and tend to overestimate rainfall frequency (seen

in Figs. 4.6b,c and Table 4.6). This corroborates results in Fig. 4.5, especially the overestimation in occurrence frequency at the low rainfall rates. The false alarms negatively influence the HSS, e.g., MSWEPv2.8 with the highest POFA has the lowest HSS, and the reverse is true for TMPA. However, the skill of SREs in detecting rainy days is better than random chance for all the products (HSS > 0). The native resolution of IMERG (IMERG-F-0.1°), as expected, has a higher HSS of 0.47 compared to the re-gridded 0.25°version (IMERG-F-0.25°) with a HSS of 0.43. This difference in skill is about 9% and may be attributed to the reduction in POFA.

Table 4.6: Summary of skill score on daily, pentadal and dekadal time steps. We do not calculate the POD, POFA, BID and HSS for pentadal, dekadal accumulations as these metrics are less meaningful at these timescales compared to a daily timescale. Similar to Tables 4 and 5, the bold number denote the best scores amongst the final products. Taken from Ageet et al. (2022). © American Meteorological Society. Used with permission.

Timestep	Satellite Product	POD	POFA	BID	HSS	r	E	ME (mm d^{-1})	PB (%)	MAE (mm d^{-1})	RMSE (mm d^{-1})
Daily	IMERG-F-0.1°	0.88	0.40	1.47	0.47	0.41	-0.04	-1.6	-16	7.6	12.4
	IMERG-F-0.25°	0.93	0.44	1.65	0.43	**0.43**	0.06	**-1.7**	**-18**	7.3	**11.8**
	IMERG-E	0.94	0.45	1.72	0.40	v0.41	-0.05	-0.1	-1	8.0	12.5
	TMPA	0.76	**0.36**	**1.19**	**0.47**	0.35	-0.13	-2.3	-24	8.1	12.9
	TMPA-RT	0.75	0.39	1.19	0.47	0.35	-0.07	-3.3	-35	7.8	12.6
	CHIRPS	0.77	0.41	1.33	0.41	0.22	-0.14	-3.1	-33	8.2	13.0
	MSWEPv2.2	0.88	0.40	1.47	**0.47**	**0.43**	0.02	-2.6	-27	7.0	12.0
	MSWEPv2.8	**0.97**	0.50	1.97	0.29	0.37	**0.07**	-3.3	-35	**7.0**	**11.8**
Pentadal	IMERG-F-0.1°					0.58	0.25	+0.6	1	14.1	20.6
	IMERG-F-0.25°					**0.60**	**0.31**	**+0.1**	**1**	13.6	9.7
	IMERG-E					0.56	0.04	+5.2	23	16.2	23.3
	TMPA					0.53	0.16	-2.0	-8	14.9	21.8
	TMPA-RT					0.52	0.15	-5.0	-22	14.6	21.9
	CHIRPS					0.50	0.19	-1.4	-6	14.8	21.3
	MSWEPv2.2					**0.60**	0.29	-2.4	-11	**12.8**	20.0
	MSWEPv2.8					0.58	**0.31**	-2.8	-12	13.3	**19.3**
Dekadal	IMERG-F-0.1°					0.68	0.42	+1.2	3	19.6	27.8
	IMERG-F-0.25°					**0.69**	**0.46**	+1.2	3	19.0	**26.9**
	IMERG-E					0.64	0.13	+10.4	26	24.0	34.0
	TMPA					0.63	0.34	-2.5	-6	20.7	29.6
	TMPA-RT					0.61	0.30	-7.9	-20	21.0	30.7
	CHIRPS					0.63	0.38	**-1.1**	**-3**	20.4	28.9
	MSWEPv2.2					**0.69**	0.45	-3.6	-9	**17.8**	27.0
	MSWEPv2.8					0.67	0.45	-3.5	-9	18.8	27.1

Considering daily rain rates, the r values suggest a low to average agreement between the satellites and gauge data (0.22 in CHIRPS to 0.43 in IMERG-F and MSWEPv2.2), while the E values show the satellite products being comparable to (IMERG-F, MSWEPv2.2 and MSWEPv2.8) or worse than (TMPA and CHIRPS) climatology. All the products underestimate the rainfall rates by 1.6 mm (IMERG-F) to 3.3 mm (MSWEPv2.8). The fact that the underestimation is mostly in the higher rain rates is confirmed by the high RMSE values (in the order of 12 mm day^{-1}) for all the products. Overall, IMERG-F emerges as the best product at a daily timescale followed by MSWEPv2.2 and TMPA, CHIRPS and then MSWEPv2.8.

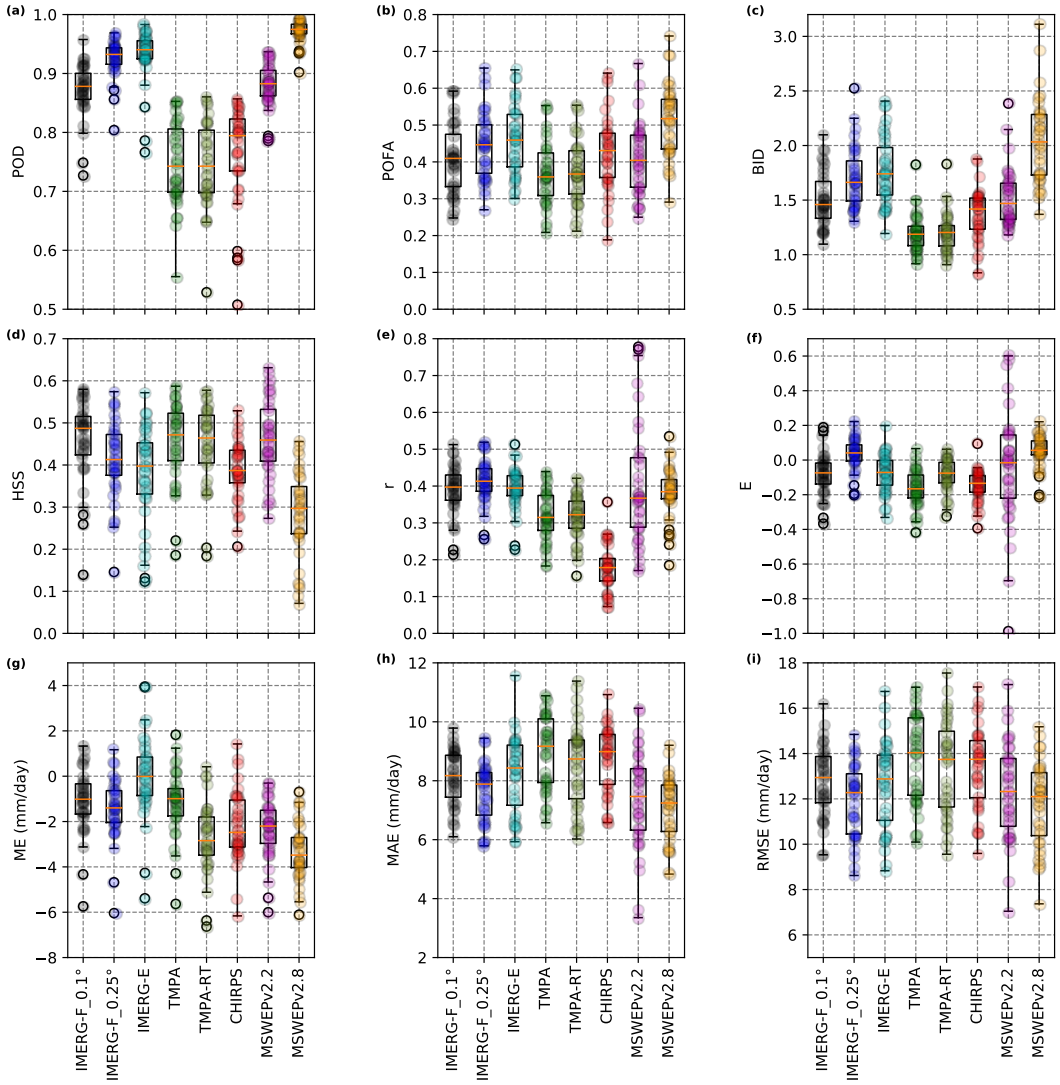


Figure 4.6: Box plots showing the performance of the satellite products on a daily timescale across all the stations. Each circle represents a station. The box is equivalent to the interquartile range (IQR) while orange horizontal line is the median (50th percentile). The lower and upper whiskers are equal to and (0.35th and 99.65th percentiles), respectively. Circles beyond the whiskers are outliers and constitute 0.7% of the data. Taken from Ageet et al. (2022). © American Meteorological Society. Used with permission.

The performance of the satellite products significantly varies across the stations (Fig. 4.6) for a variety of reasons. One explanation is the varying weights applied in the gauge-calibration of the SREs. The comparison of two subsets of stations: a) GTS (7 stations whose data is reported to the GTS and have >49% availability of daily data, Fig. 4.2a), and b) non-GTS (stations which do not report to the GTS, hence, were not used in the gauge-calibration) confirmed this assumption. The QQ plots (Fig. A.4) show that in all the SREs, the distribution of rain rates for the non-GTS station differs more substantially from that of the RGs compared to the GTS stations. On average,

the performance of the SREs at the pixels closest to the GTS stations is better than that at the pixels closest to the non-GTS stations for all the metrics (Fig. 4.7). The standard deviation (light-red shading in Fig. 4.7) demonstrates that for any combination of non-GTS stations, the performance of the GTS stations is always superior. It is further argued here that the overall highest station-to-station variability observed in most of the non-dichotomous (i.e., r, E, ME, MAE and RMSE) metrics for MSWEPv2.2 is most likely related to the strong weights of gauge observations in this product at grid points close to the GTS stations. MSWEPv2.2 showed the highest spatial variability for r, E, MAE and RMSE while IMERG-E showed the highest variability for ME (Fig. 4.6e–l). The disparities in performance at the GTS stations and non-GTS stations were also largest in MSWEPv2.2. For example, based on Figure 4.7, there was a difference of 0.35, 0.5 and 0.21 for the r, E and HSS, respectively, in MSWEPv2.2 compared to, for example, 0.11, 0.15 and 0.08 in IMERG-F for the same metrics. MSWEPv2.8 is also substantially different from its predecessor, with smaller differences of 0.12, 0.17 and 0.06, respectively. The difference between the two versions of MSWEP may partly be explained by the reduction in gauges used in the newer version due to stricter gauge inclusion criteria (Beck et al., 021b).

The performance of the SREs in the different cluster varies (Fig. 4.8). The POD is highest in MWSEPv2.8 (Fig. 4.8a), but the skill is degraded by the high rate of false alarms (Fig. 4.8b). The correlation coefficient (Fig. 4.8d) is best in IMERG products, followed by the MSWEP products. However, as seen previously, the variability in performance is very large in MSWEPv2.2. The C- and SW-clusters exhibit correlation values of about 0.32 while that of the LVB-cluster is substantially higher at 0.65 (Fig. 4.8d). This superiority in skill at the LVB-cluster is partly due to the fact that majority of stations in this cluster are GTS stations. Products perform worst in SW-cluster with the highest rate of false alarms and hence, lowest HSS for all the products (Figs. 4.8b,c).

As expected, aggregating daily rainfall totals to longer timescales improves the skill in all the SREs (Table 4.6 and Fig. 4.9), since any errors due to time mismatches are reduced. The r and E values of all the products improve considerably. For example, the correlation value in CHIRPS increases from 0.22 to 0.50 moving from daily to pentadal accumulations (Table 4.6). This is likely because CHIRPS is initially created at pentadal timescale and then disaggregated into daily data (Funk et al., 2015). Generally, the performance of the satellite products converges with longer timescale. The “Taylor score” Taylor (2001), which combines the performance with respect to correlation and standard deviation, increases (i.e., improves) with increasing timescale (Fig. 4.9). IMERG products overestimate both pentadal and dekadal rainfall while the other SREs underestimate rainfall at these time scales. All the products underestimate the variability and amount of rainfall at a daily timescale (all points below the RG standard deviation and the arrows point towards the centre, Fig. 4.9). IMERG-F performs best

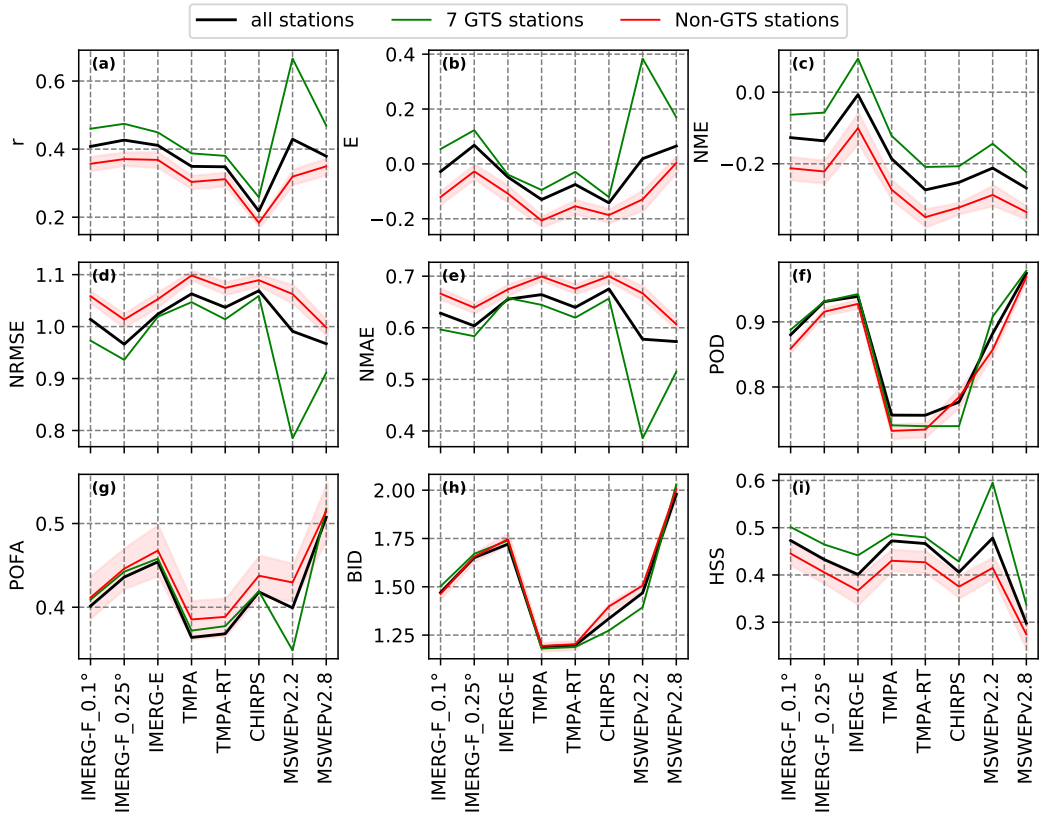


Figure 4.7: Metrics showing the performance of 7 stations with the highest percentage of data reported to the GTS versus performance of station which never report their data to the GTS (non-GTS). The red curve is the ensemble mean of all the possible 7 stations out of the 12 synoptic and the light red shading is the standard deviation. The black line shows the performance of all the stations. Note that the normalized version of ME, MAE and RMSE are preferred here because the station subsets being compared are different. Taken from Ageet et al. (2022). © American Meteorological Society. Used with permission.

at pentadal and dekadal timesteps for most of the metrics, followed by MSWEPv2.2 and MSWEPv2.8, CHIRPS and then TMPA.

4.2.4 Extremes and return periods

Due to variations in the rainfall climatology as seen in Sections 4.2.1 and 4.2.2, extreme events are independently determined for individual stations. First, we considered the daily rainfall totals above the 95th percentile of all rainy days at a given station as "extreme". The distribution of extreme events differs significantly between the RGs and SREs for all the SREs (Fig. 4.10). All the points are below the diagonal meaning the SREs fail to capture extreme rain rates in all the clusters. The deviation is largest in MSWEPv2.8 and CHIRPS while IMERG-F and TMPA are the closest to the 1:1 line. IMERG-E exhibits a better distribution than IMERG-F while TMPA-RT is worse than TMPA. IMERG-F-0.1°performs better than the IMERG-F-0.25°as expected since coarse

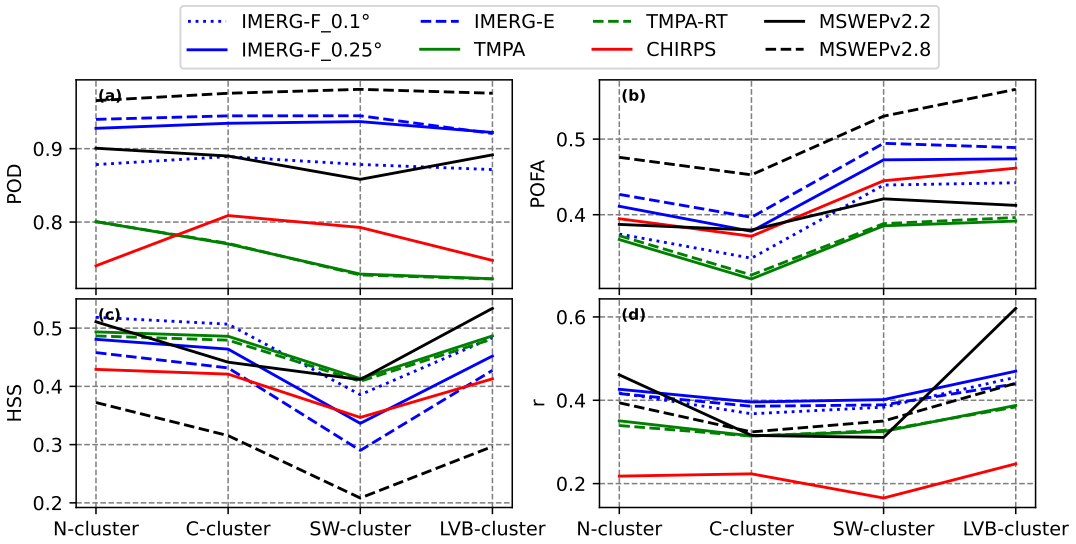


Figure 4.8: Statistics for the four clusters calculated by concatenating daily rainfall totals for all the stations in a cluster for the period 2001–2018 (IMERG-F-0.1°, IMERG-F-0.25°, IMERG-E, TMPA, TMPA-RT, CHIRPS and MSWEPv2.8) and 2001–2016 (MSWEPv2.2). All the correlations are statistically at 99% significance level. Taken from Ageet et al. (2022). © American Meteorological Society. Used with permission.

graining averages extremes out. Next, we checked the distribution of extreme events in the SREs (the 95th percentile of the individual SREs is taken as the thresholds) and the corresponding values in the RGs (Fig. A.5). In this case, the points are above the diagonal for the weaker extremes suggesting that the SREs overestimate these extremes. However, at the stronger extremes, the plot moves below the diagonal pointing towards a tendency of the SREs to underestimate the stronger extremes. Similar results were shown by Monsieurs et al. (2018).

The performance of the SREs declines for extreme events for all the satellite products at all stations in comparison to when the full dataset is used (cf. Fig. 4.6). The ability of the satellites to detect extreme events decrease to a POD < 0.21 for all final products (Table 4.7). The skill of the satellites is also reduced (lower HSS values compared to Fig. 4.6 and Table 4.6) due to the increased POFA in all the products. This is most pronounced in CHIRPS (Figs. 4.11b,d and Table 4.7). The satellites also underestimate the frequency of extreme rainfall events (BID < 1). All the products greatly underestimate the extreme rain rates (Fig. 5.10g) on the order of 26 mmd^{-1} (53%) or more (Table 4.7). The skill is worse than climatology for all the products (Fig. 4.11f and Table 4.7). The satellite estimates are weakly (0.06 - 0.26) correlated to the RGs (Fig. 4.11e and Table 4.7). Overall, for the final version of the products, MSWEPv2.2 marginally emerges as the best product for extreme events, followed by IMERG, TMPA, MSWEPv2.8 and then CHIRPS. However, as seen previously, the performance of MSWEPv2.2 is highly variable suggesting the influence of the applied gauge weighting. Comparing the final products of IMERG and TMPA with their respective early versions,

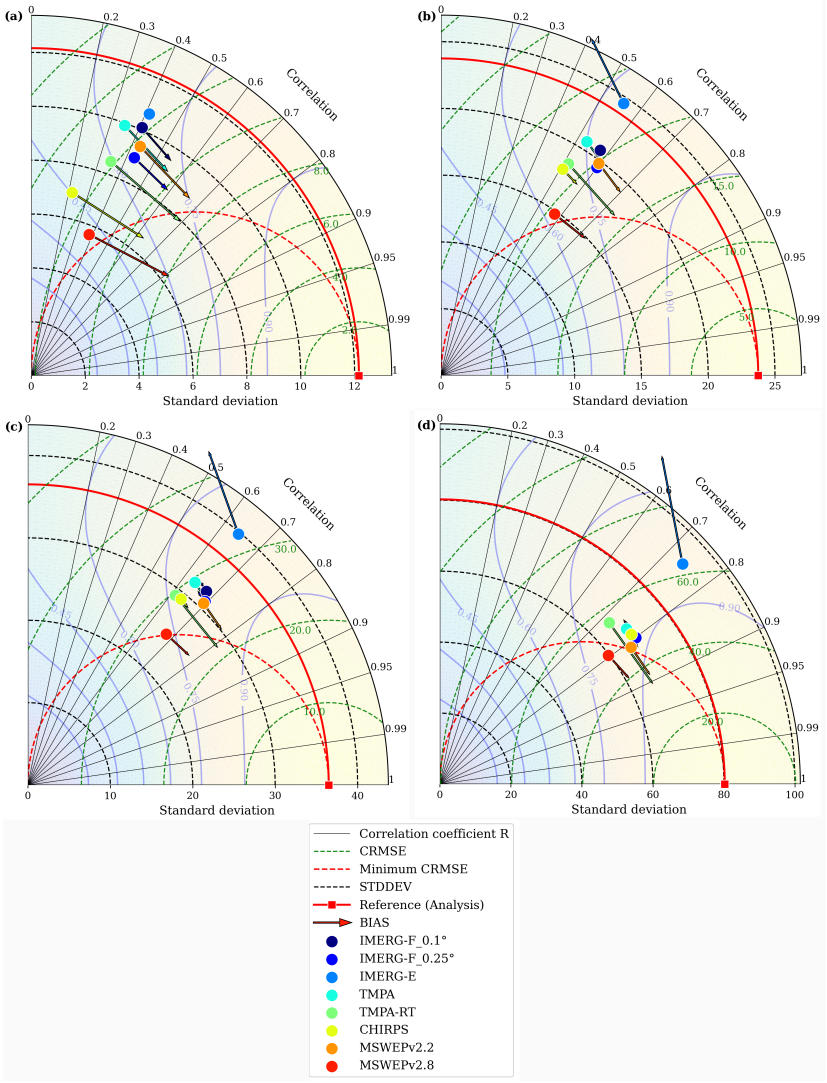


Figure 4.9: Taylor diagrams showing the correlation coefficient, standard deviation, centered RMSE and the Taylor score between gauge data and the satellite estimates on: (a) daily; (b) pentadal; (c) decadal; and (d) monthly timescales. The scores here are calculated by concatenating all the data for all the station for the individual timescales. The red square marker on the horizontal axis is the reference (gauge) dataset, and therefore, the closer a satellite product to the reference, the better it is. The length of arrows represents the bias magnitude while the direction the sign of the bias with arrows pointing towards (away from) the reference means underestimation (overestimation). The bias shares the same axis as the CRMSE, and its magnitude in mm is equivalent to the length of the arrow read off starting from the origin (red square). The shading and blue contour indicate the Taylor score (combines correlation coefficient and standard deviation, and ranges from 0 (worst) to 1(best)). Taken from Ageet et al. (2022). © American Meteorological Society. Used with permission.

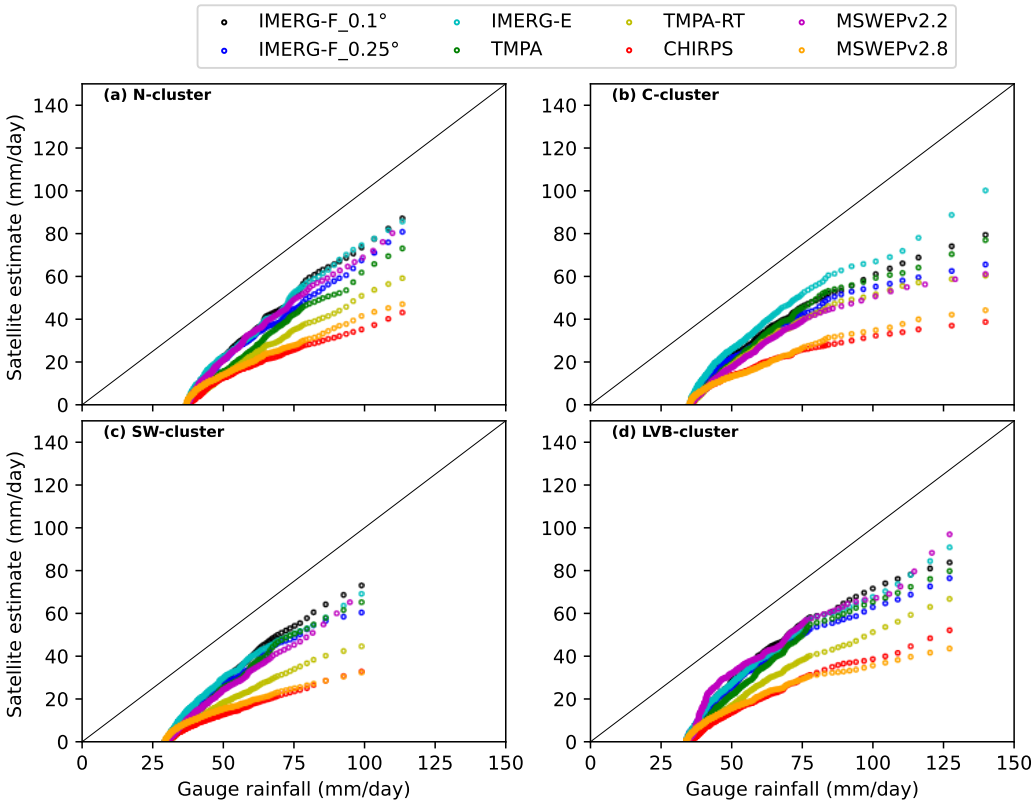


Figure 4.10: QQ plots showing the distribution of extreme rainfall as recorded by the RGs and the corresponding values in the SREs in the pixel closest to the stations. Taken from Ageet et al. (2022). © American Meteorological Society. Used with permission.

Table 4.7: Same as the section for daily accumulations in Table 4.5 but for daily extremes events (rainfall amount > 95th percentile in rainy days subset of RGs). Please note that the POD, POFA, BID and HSS are computed based on the extreme threshold for the individual stations and this threshold is applied to the whole rainy-day (>0.2 mm) subset. The rest of the metrics consider the extremes in the RGs and the corresponding SREs values. The numbers in bold and italics font denote the best scores amongst the final versions and scores when an early product outperform all products, respectively. Taken from Ageet et al. (2022). © American Meteorological Society. Used with permission.

Satellite Product	POD	POFA	BID	HSS	r	E	ME (mm d^{-1})	PB (%)	MAE (mm d^{-1})	RMSE (mm d^{-1})
IMERG-F-0.1°	0.18	0.70	0.59	0.19	0.18	-4.28	-28.9	-58	30.8	35.6
IMERG-F-0.25°	0.15	0.68	0.43	0.18	0.19	**-4.28**	-29.8	-61	30.9	**35.5**
IMERG-E	0.21	0.72	0.76	0.21	0.16	-3.78	-26.5	-53	29.1	34.1
TMPA	0.14	0.75	**0.56**	0.15	0.14	-5.07	-32.4	-62	34.0	38.6
TMPA-RT	0.10	0.72	0.37	0.13	0.14	-5.50	-35.3	-68	36.0	40.4
CHIRPS	0.02	0.81	0.12	0.03	0.06	-6.30	-38.7	-77	38.8	42.5
MSWEPv2.2	**0.19**	0.56	0.42	**0.24**	**0.26**	-4.32	**-29.5**	**-59**	**30.5**	35.7
MSWEPv2.8	0.03	**0.55**	0.07	0.05	0.15	-5.61	-36.2	-74	36.3	39.8

IMERG-E is better than IMERG while TMPA-RT is worse than TMPA for extreme events detection and accuracy. Indeed, considering all the SREs, IMERG-E emerges best overall for extreme events, having the best scores in all but one metric evaluating rain rates, the best extreme-events detection percentage (21%), and the lowest BID (Table 4.7). This suggests that IMERG-E is best suited for extreme events analysis.

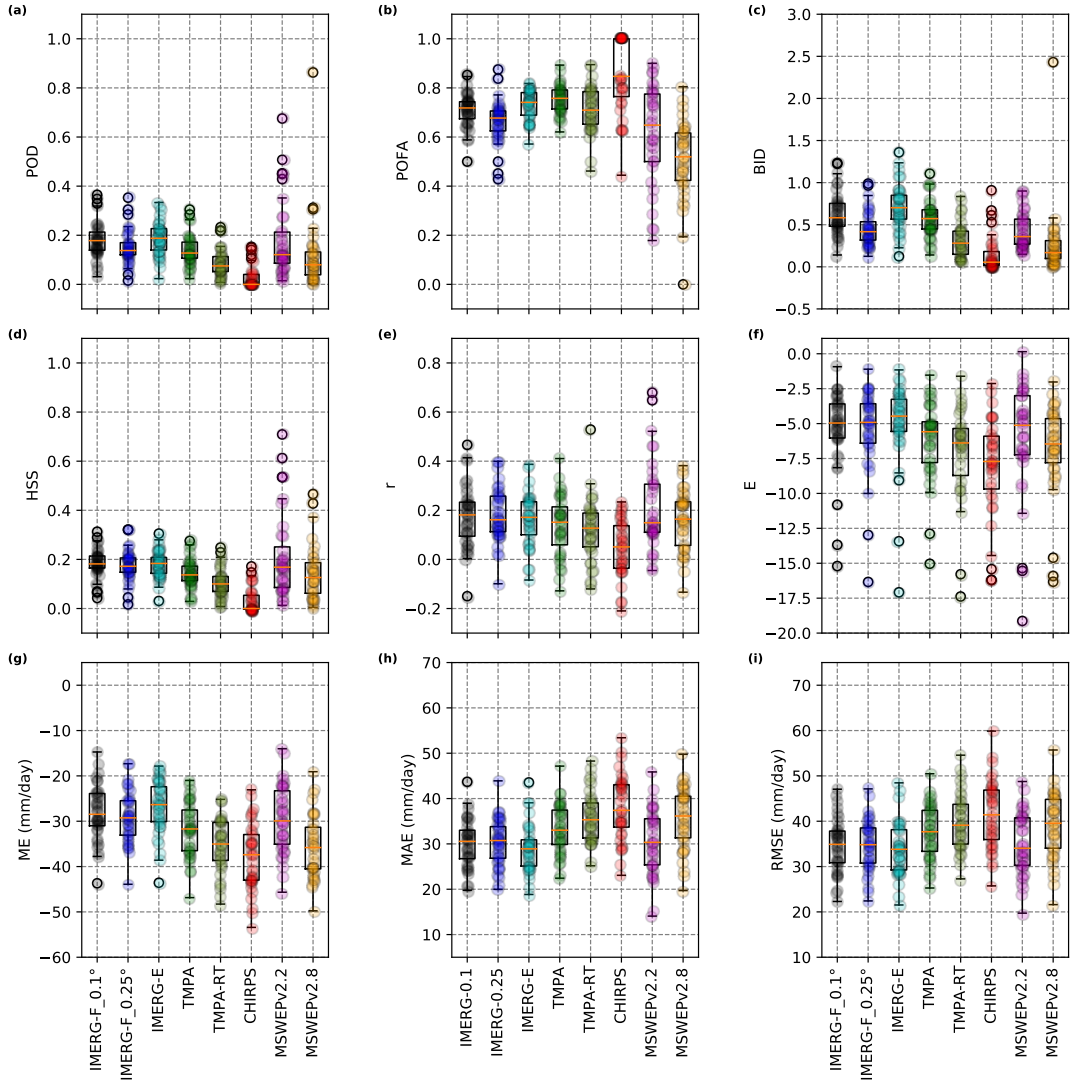


Figure 4.11: Same as in Figure 4.7 but for extreme events. A hit occurs when both the RG and SRE record rainfall exceeding the 95th percentile of rain-days subset of the RG data. Taken from Ageet et al. (2022). © American Meteorological Society. Used with permission.

With regard to rare events, modelled with the POT method, the performance of SREs is spatially variable and the products generally underestimate the return values (Fig. 4.12 and Fig. A.6). Note that the return values are dependent on length and quality of available data (results not shown). Therefore, we constrained the analysis to the period covered by all products (2001–2018), except for MSWEPv2.2 (2001- 2016). Figure 4.12 illustrates that SREs perform satisfactorily at some stations, e.g., in Kitale (Fig. 4.12a)

for MSWEPv2.2 and poorly in other stations. Apart from nine stations in MSWEPv2.2, all SREs underestimate the return values at all other stations (see Fig. A.6). The underestimation is most pronounced in CHIRPS, MSWEPv2.8 and TMPA-RT.

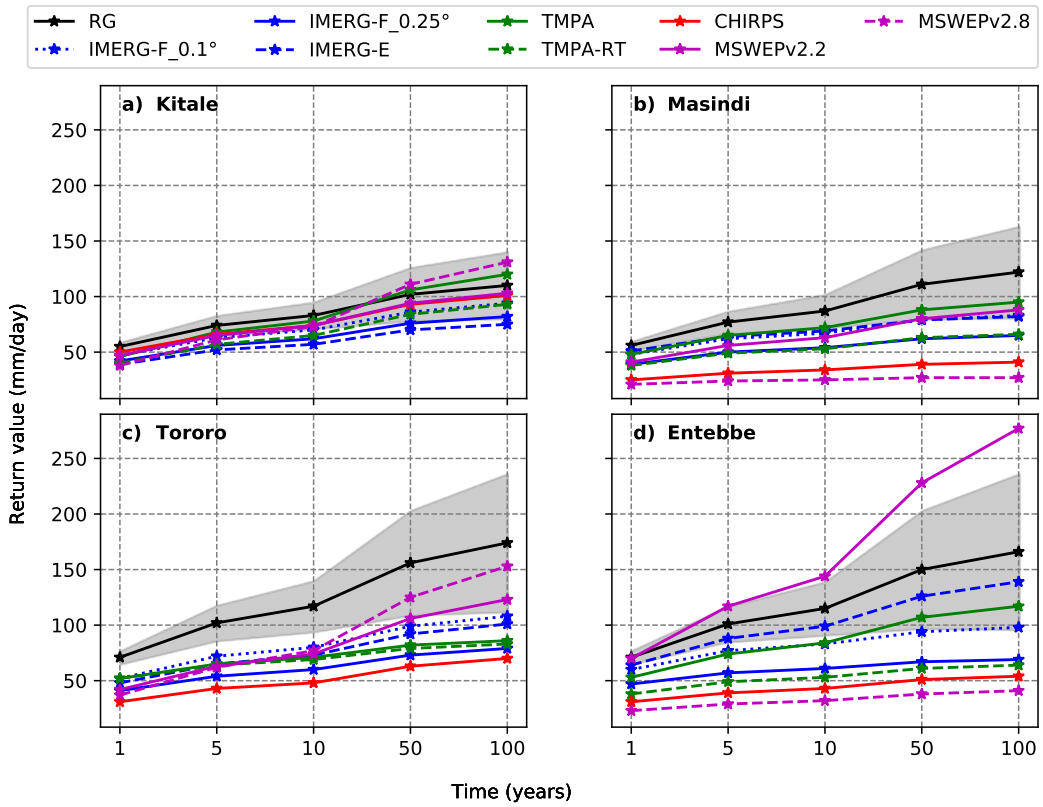


Figure 4.12: Return values based on the POT method RG and the SREs for four randomly select stations in the study area. The grey shading is the confidence bounds of the return values modelled using RG data. Taken from Ageet et al. (2022). © American Meteorological Society. Used with permission.

Since this study has a strong focus on comparison between SREs, we converted the absolute return values at each station to relative return value deviations in order to make the return values comparable across the stations in different climates and datasets. Considering the mean values, all the products always underestimate the return values of extreme events except MSWEPv2.2 which overestimates after about 40 years (Fig. 4.13). The underestimation and overestimation of the return values at the shorter and longer periods, respectively, coupled with a large standard deviation especially at long return periods (Fig. 4.13h) does not allow for a conclusive statement about the general usefulness of MSWEPv2.2 for the analysis of extreme rainfall. The performance of IMERG is greatly affected by the coarse graining, since the native resolution performs best amongst all products, but the coarse-grained version is only better than TMPA-RT, MSWEPv2.8 and CHIRPS. The early product of IMERG also outperforms the final product.

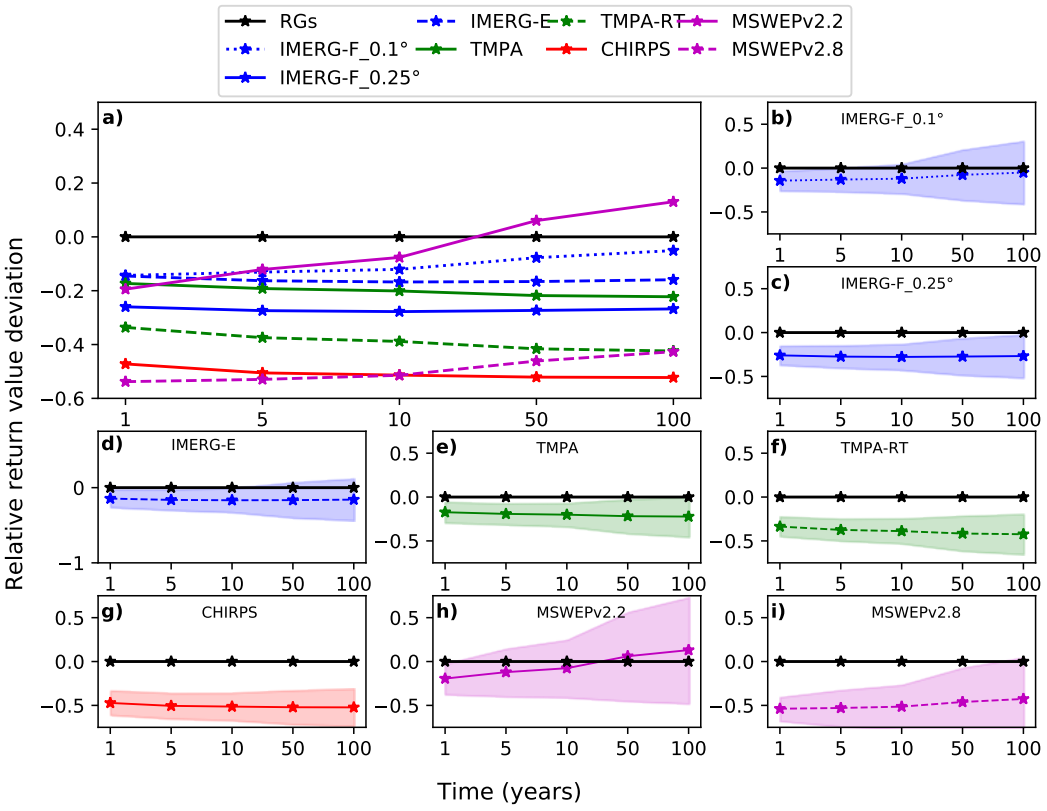


Figure 4.13: Modelled relative return values of extreme events for all the satellite estimates (a) and, b–h shows individual satellite product relative return value plus the standard deviation across stations (shaded). For each dataset, the values are obtained by taking the mean across all stations and time. Taken from Ageet et al. (2022). © American Meteorological Society. Used with permission.

4.3 Discussion and conclusions

We analysed the performance of four satellite products against daily RG data over the complex topography of Uganda at multiple spatio-temporal aggregations for the period 2001–2018 (except MSWEPv2.2, 2001–2016). Additionally, we compared the early versions of IMERG and TMPA, both against each other and with their respective final versions. The native resolution of IMERG was also validated to evaluate how resolution affects the performance of satellite products. We included the two versions of MSWEP in order to compare the change in performance of the latest version compared to its predecessor. Since RG data was missing for some periods, we used satellite data only at corresponding timesteps with available RG data. Two sets of metrics that assess rainfall occurrence and accuracy of the rainfall totals were used. The suitability of the products around extreme events was also tested using the same metrics. Additionally, using the POT method, we assessed if the SREs can be used to simulate return values of rare events for specified return periods. The main conclusions of this validation study are:

1. Generally, all the products reproduce the annual and seasonal rainfall pattern, but the amounts are overestimated in IMERG by about 5% and underestimated by TMPA, CHIRPS and MSWEP by between 5–8% (annually) and 1–9% (monthly).

2. IMERG outperforms the other products on daily, pentadal and dekadal timescales followed by MSWEPv2.2, TMPA, and lastly CHIRPS. On monthly and annual timescales, MSWEPv2.2 and CHIRPS, respectively, perform best. MSWEPv2.8 performs worse than its predecessor, MSWEPv2.2 for almost all the spatio-temporal aggregations and for extremes.

3. The performance of the satellites is greatly reduced for southwestern part of Uganda, mainly due to the high false alarm rate which degrades the skill.

4. The performance of the all the SREs is spatially variable for reasons mostly elusive to the user. However, MSWEPv2.2 is a salient example where a high gauge weight is capable of drastically improving the performance at gridpoints close to the station, but the skill drops considerably in places away from the RGs used in the calibration. Unless documented thoroughly, this can lead to serious non-transparency of a SRE's skill for the end-user.

5. Except for IMERG-E, all the products miss more than 80% of the daily extreme events reported by the RGs and severely underestimate extreme daily rainfall totals. The fact that IMERG-E outperform all the other products demonstrates the potential of NRT products for operational applications such as floods assessment, which need short latency products.

6. IMERG, IMERG-E, TMPA and MSWEPv2.2 show closest agreement with the RGs for return values analysis while MSWEPv2.8, CHIRPS and TMPA-RT should be avoided for this purpose. This is not surprising given that the SREs that are the least suitable for modelling extreme events underestimate extreme rainfall events most severely.

Good agreement of the products with RGs on a seasonal scale supports the findings in Camberlin et al. (019a), Diem et al. (2014) and Asadullah et al. (2008) with similar performance statistics and seasonal cycles which are largely controlled by the periodic migration of the tropical rain belt Nicholson (2017); Seregina et al. (2018). Therefore, SREs capture the drivers of synoptic seasonal rainfall over the study area well. The fact that all the products used in the present study are gauge calibrated may have contributed to the good performance (Awange et al., 2015; Dinku et al., 2018). Additionally, at longer temporal accumulations, any errors due to time mismatches between the RG and SREs data are reduced, which, in turn, improve the scores.

The superior skill of IMERG compared to the other SREs, especially at lower temporal aggregations has been seen in other studies. For instance, Dezfuli et al. (017a), showed that IMERG was better than TMPA in West and East Africa. In general, the good

performance of IMERG can be explained by better performance of PMW sensors (Kidd and Huffman, 2011; Bitew and Gebremichael, 2011) and improved spatio-temporal resolution (Dezfuli et al., 017b). However, IMERG overestimates rainfall occurrence and intensity compared to both the RGs and other SREs, especially over Lake Victoria. This is attributed to the tendency of PMW sensor to overestimate rainfall in convectively active regions (O and Kirstetter, 2018; Nicholson et al., 2021). The generally good performance of MSWEPv2.2 agrees with those of Lakew et al. (2020). However, MSWEPv2.8 has a weaker performance compared to MSWEPv2.2. This may stem from; a) the assignment of large weights to ERA5, which has a tendency to rain too often at low rain rates, and b) the reduction in the number of RG data used in gauge-calibration due to a more stringent selection criterion (Beck et al., 021b). The superior performance of CHIRPS compared to the other SREs at larger temporal accumulations corroborates the findings of Dinku et al. (2018), Diem et al. (2019), and Camberlin et al. (019a) and may be due to; (a) use of the gauge-satellite climatology, CHPclim which reduces mean biases, and (b) the fact that gauge calibration is done at longer temporal resolutions (i.e., pentadal and monthly).

SREs are known to show variable performances depending on different factors. Although other factors like the gauge calibration and topography (e.g., Diem et al., 2014; Monsieurs et al., 2018) play a significant role in the variability of the performance of SREs in the study area, we specifically found for MSWEPv2.2 that the weight assigned to RGs relative to the other data inputs during gauge calibration process had a significant influence on how the product performed at a given grid point. This result corroborates several studies (e.g., Awange et al., 2015; Camberlin et al., 019a) who also showed that gauge calibration impacts the performance of SREs.

Major difficulties in capturing rainfall by SREs in the domain of the SW-cluster have already been highlighted in Diem et al. (2014) and Monsieurs et al. (2018). The complex topography partly accounts for the performance. Many parts of the region lie in a rain shadow which causes false alarms. Additionally, compared to the other three clusters, the SW-cluster is drier, hence, sub-cloud evaporation could be more frequent in this region. Sub-cloud evaporation increases cases false alarms, and this has been previously shown by Dinku et al. (010a) over the desert locust regions and Thiemig et al. (2012) over the semi-arid Juba-Shabelle region.

Similar to studies by Monsieurs et al. (2018) over Western Uganda and Thiemig et al. (2012) over African river basins, the performance for all the SREs declines for extreme events compared to the whole dataset. This may be partly due to the fact that PMW sensors may miss the short-lived intense rainfall events given their infrequent overpasses. Additionally, being gridded products, SREs contain spatial averages and hence, extremes may be smoothed out, which is not the case for the point measurements at RGs. The better detection rate and reduced error in IMERG-E around extreme

events compared to the other SREs points to the fact that gauge calibration introduces some uncertainties in the final version of the SREs (Bitew and Gebremichael, 2011). For most SREs, the calibration is initially done at a low temporal resolution, e.g., monthly for IMERG, before daily rainfall is rescaled accordingly. This in itself may be problematic for daily extremes. However, for the NRT product of TMPA, results showed overall poorer scores compared to the final version. The stronger underestimation of TMPA-RT was also found by Monsieurs et al. (2018) and this was attributed to the lack of gauge calibration. Unsurprisingly, the SREs with highest errors at the extreme events (CHIRPS, TMPA-RT and MSWEPv2.8) had the highest error compared to the RGs when applied for modelling the return period of extreme events.

Based on this study, care needs to be taken when using the SREs for a given application in the studies as performances varied substantially. This variation stems from various sources, e.g., gauge calibration, algorithms, data inputs and region considered. We are also aware that the performances could also have been influenced by the availability and quality of RG dataset. Also, given that the SREs ranking was different for individual metrics, it is difficult to categorically say which product is the best. The SREs showed great promise at the longer temporal aggregations given the high scores in most metrics. Challenges remain for shorter time scales, especially over heterogeneous topography, and extreme rainfall events. Despite these challenges, IMERG, TMPA and MSWEPv2.2 are the most suitable products among the tested SREs for hydro-meteorological disaster applications as they provide the crucial information for a data sparse region like Uganda. Validating the SREs with a denser gauge network may highlight the apparent regional difference in performance even further and also enable the quantification of the influence of the RGs on the results, if any.

5. Skill of ECMWF reforecasts over EEA

In this chapter, the analysis of the ECMWF reforecast quality is presented. Like in the previous chapter, the data sets and methods are briefly introduced in the first section. Then results of the assessment of the skill of both the deterministic and probabilistic forecasts are presented in section 5.2. The analysis was also done at regional and whole domain spatial considerations and for 24-, 48- and 120-hour accumulations. The improvement in skill after postprocessed the raw reforecasts are also presented. Finally, in section 5.3, the discussion and conclusions are presented.

5.1 Data and Methods

5.1.1 Data

5.1.1.1 ECMWF rainfall hindcasts

The study uses precipitation reforecasts from the ECMWF model available from the S2S prediction project database (Vitart et al., 2017). The reforecasts are generated on the fly with respect to the near-real-time forecasts, initialized twice a week, have a lead-time of 46 days, and spatial (temporal) resolution of 1.5° (6-hourly). The reforecats used here were generated using the model version dates from 2020, which were based on CY46R1 and CY47R1 cycles of the Integrated Forecasting System (IFS) model. For the 20-year period, 2000–2019, we have 105 initialization dates per year resulting from 2 forecasts in each of the 52.5 weeks. The reforecast is made up of 11 ensemble members (1 control and 10 perturbed).

5.1.1.2 Satellite rainfall estimates

The first set of precipitation observations used for the verification of the reforecasts are satellite rainfall estimates, specifically, the final daily product of Integrated Multisatellite Retrieval for Global Precipitation Measurement (GPM) v6B (IMERG; Huffman et al., 2020). Being a satellite-based product, IMERG was chosen because it offers a more complete spatio-temporal coverage for the data sparse regions like EEA (Diem et al., 2014; Dinku, 2019) compared to ground-based observation, e.g., rain gauges or radars. Moreover, it has been shown to be among the best performing products for this region

at daily to monthly time scales (Ageet et al., 2022). IMERG is available from June 2000 to date at a temporal resolution of 30 minutes and a spatial resolution of 0.1°. Since the reforecasts are available at a spatio-resolution of 1.5°, IMERG rainfall estimates are regridded to the same resolution using first-order conservative remapping (Jones, 1999).

5.1.1.3 Rain gauge rainfall

Additionally, in-situ rainfall observations from rain gauges in Uganda available from the Karlsruhe African Surface Station-Database (KASS-D; Vogel et al., 2018) are used. Because of the relatively coarse resolution of the gridded products, the analysis with gauges was done at nearest grid-points, with a requirement that the grid-point is nearest to at least four stations. Additionally, the stations considered should have at least 95% daily data availability in the period of study. Three grid-points satisfied all the above conditions. Due to their location, we named these grid-points, Lake Victoria, Lake Kyoga and western Uganda regions (Fig. 1.1). The period 2001–2018 offered the most complete record and was therefore used for the analysis. The mean of the stations at a grid-point was used for the validation of the forecasts. Because the daily gauge rainfall in this region is accumulated from 06 UTC for 24 hours until 06 UTC of the following day. The reforecast and IMERG rainfall were aggregated to match the gauges accumulation period and all the analyses was done for the common period of 2001–2018.

5.1.2 Methods

5.1.2.1 Verification methods

Forecast verification primarily assesses how 'well' a forecasting system predicts the target variable based on the observed values. Because the goodness of a forecast may depend on more than one attribute and the purpose of the verification (Murphy, 1993), a single verification method is not sufficient to assess the predictive performance of the forecast. Hence, several methods should be taken into account. To assess different aspects of the ECMWF hindcasts, we assess the ensemble both in probabilistic and deterministic terms, the latter in form of the ensemble median. While the ensemble is used to assess the predictive performance of the hindcasts in terms of scores, calibration and discrimination, the ensemble median is used to assess the accuracy and bias of a point forecast derived from the ensemble, which we refer to as a deterministic forecast. Further, we derive probability forecasts for the occurrence of rainfall from the ensemble by calculating the fraction of ensemble members that predicts rainfall. The different metrics are briefly explained below.

For the deterministic forecast, the mean error (ME) is computed as the mean difference between the ensemble median of the hindcasts and the observations, with positive (negative) values indicating overestimation (underestimation). Additionally, the mean absolute error (MAE) is computed for different rainy day thresholds (i.e., when the

observations have a precipitation accumulation of more than a given threshold value), in this case from 0 up to 10 mm, to show how the error changes for higher rain rates. If f_i and o_i are the point forecast and observation at time i, and N is the sample size, then the metrics above are defined as:

$$ME = \frac{1}{N}\sum_{i=1}^{N}(f_i - o_i), \tag{5.1}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|f_i - o_i|. \tag{5.2}$$

To assess the discrimination ability or the potential prediction ability of probability forecasts for occurrence of rainfall, the receiver operating characteristic (ROC) curve and area under the ROC curve (AUC) are suitable tools (Wilks, 2011). The ROC curve is generated by plotting the hit rate against the false alarm rate at different probability thresholds. The hindcasts have no discrimination ability if the curve falls on the diagonal and a perfect discrimination is obtained if the curve passes at the top-left corner. The area under the diagonal and the curve gives the AUC, with values between 0 (no discrimination) and 1 (perfect discrimination). By comparing to a reference, EPC (EPC15, hereafter, as observations in a ±15 days-window around the date of interest are considered), we compute the the so-called AUC skill (AUCS) like in Walz et al. (2021). For every forecast date, we generated the EPC15 by taking past observations on this date and the 30 days around it, yielding an ensemble of 527 members (i.e., 31 members times 17 years). Details and code to compute the EPC is given in Walz et al. (2021).

Most probabilistic forecasts do not quantify the forecast uncertainty adequately meaning they are miscalibrated or unreliable (Wilks, 2011). To check the reforecasts for calibration, we use standard tools from forecasting methodology (Gneiting and Katzfuss, 2014). Rank histograms are used to assess the calibration of the ensemble forecasts, probability integral transform (PIT) histograms are used for the postprocessed forecasts and EPC15. Note that the PIT histograms for EPC15 are calculated as described in Vogel et al. (2018), Vogel et al. (2020) and Schulz and Lerch (2022). Both rank and PIT histograms can be interpreted analogously, where a flat histogram corresponding to a uniform distribution indicates that the forecast are calibrated while a U-shaped (hump-shaped) histogram indicates underdispersed (overdispersed) hindcasts, that is, the forecasts are overconfident (underconfident). The calibration of probability forecasts is checked via reliability diagrams, which show the calibration curve that plots the conditional event probability of the dichotomous event against the associated forecast probabilities. If the curve is close to the diagonal, the forecast is said to be calibrated or reliable. Here, we use a new approach which ensures Consistency, Optimality, Reproducibility and is based on the pool-adjacent-violators (PAV) algorithm (CORP; Dimitriadis et al., 2021).

A major advantage of this approach is that it generates 'optimally binned, reproducible, and statistically consistent reliability diagrams' (Dimitriadis et al., 2021).

A quantitative evaluation of the forecast performance is done using proper scoring rules, which yield the best scores (in expectation) when we forecast the 'true' underlying distribution of the observation (Gneiting and Raftery, 2007; Wilks, 2011). For assessing probability forecasts of rainfall occurrence, the most common metric is the Brier score (Brier, 1950), while for the rainfall amounts, the continuous rank probability score (Gneiting and Raftery, 2007) is used. The mean Brier score (BS) for a sample of size N is defined as

$$BS = \frac{1}{N}\sum_{i=1}^{N}(p_i - y_i)^2, \tag{5.3}$$

where p_i is the probability forecast at time i and y_i the corresponding observation, which takes the value 1 for rainfall occurrence and 0 elsewise. If F_i is the cumulative distribution function (CDF) of a precipitation forecast at time i, and o_i is the corresponding observation, the mean continuous rank probability score (CRPS) of a sample of size N is defined as

$$\mathrm{CRPS} = \frac{1}{N}\sum_{i=1}^{N}\int_{-\infty}^{\infty}\left[F_i\left(x\right) - 1\left(x \geq o_i\right)\right]^2 \mathrm{d}x. \tag{5.4}$$

Both scores are negatively oriented meaning that smaller values indicate superior predictive performance. The CRPS is in the unit of the observation, mm, in this case. For scores of a method, the corresponding skill scores, i.e., Brier skill score (BSS) and continuous rank probability skill score (CRPSS) were computed relative to EPC15. Negative skill scores indicate that the method performs worse than the reference forecast, a skill score 0 that equal performance and a positive skill score that the method outperforms the reference with 1 corresponding to perfectly forecasting the observed values. For detailed discussion of these, and the other metrics used in this manuscript, we refer the reader to Schulz and Lerch (2022).

To investigate the sources of the strengths and weaknesses of the forecast, we use the CORP approach to decompose the BS in a miscalibration (MSC), discrimination (DSC) and uncertainty (UNC) component. MCB quantifies the degree of miscalibration of the forecasts (smaller values are preferred), while DSC evaluates the ability to discern between events and non-events (larger values are better), and finally UNC which is purely based on the observation indicates how difficult the forecasting problem is. To check if any observed differences in performance between the reforecasts and EPC15 are significant, we applied the (2-sided) Diebold-Mariano test (DM test; Diebold and Mariano, 1995) to forecast-EPC15 score pairs (BS or CRPS) at each grid-point. The DM test checks the hypothesis whether the (raw or postprocessed) reforecasts and EPC15 have the same expected score, that is, equal predictive performance. Because we are testing multiple grid-points independently, there is need to account for possibility of the false discovery rate (Wilks, 2016). We therefore applied a Benjamini and Hochberg

(1995) procedure which controls the proportion of falsely rejected hypothesis at a chosen significance level, 0.05 in this case.

5.1.2.2 Postprocessing

In order to remove or reduce systematic errors of the ensemble forecasts such as biases and dispersion errors, methods from statistical postprocessing are typically used to correct for them (see, e.g., Gneiting et al., 2005; Vannitsem et al., 2018; Schulz and Lerch, 2022). Here, we apply the non-parametric isotonic distributional regression (IDR; Henzi et al., 2021) method. The method assumes an isotonic relationship (i.e., an increase of the predictor variable yields an increase of the predictand) between the forecasts and observation. In this case the ensemble members were used as the predictors, following the componentwise partial order. The advantages of this approach over other postprocessing methods such as EMOS are that; 1) it is a generic method that can be applied directly as it does not require any prior conditioning or tuning, and 2) it estimates a flexible data-driven forecast distribution based only on the assumption of isotonicity. The method has also been shown to be work well in other postprocessing applications (e.g., Maier-Gerber et al., 2021; Schulz and Lerch, 2022).

For the postprocessing, the data were divided into a training and test set. As noted by Henzi et al. (2021), IDR being a non-parametric method, requires quite a large training period for the model to sufficiently learn the forecast-observation relationship. We therefore divided the data into 17 years of training, and one year for testing. Although we use a local approach, whereby we train and apply the the model for each grid-point separately, data from the eight grid-points surrounding the grid-point of interest are incorporated for training the model. This increased the size of the training data set and is reasonable, given the similarity in rainfall climatology of neighboring grid-points.

5.1.2.3 Spatio-temporal considerations

Because the performance of the forecast varies with lead-time and is influenced by the underlying topography, we assess the forecasts at different temporal aggregations and in four different locations in the region (cf. Fig. 1.1) and for different temporal aggregations. The different regions are; a) Uganda, characterized by a mixture of mountains, large water bodies and flat land, b) the Congo Basin, a vast area of mainly low-lying and forested region, 3) East African Coast, a coastal region along the shores of the Indian Ocean, and 4) Ethiopian highlands. For the temporal aggregations, we consider 24-hour, 48-hour and 120-hour accumulations, and increasing lead-time, that is, 1–14 days ahead. The longer aggregations are important to certain economic sectors. For example, to farmers, the exact timing of rain may not be so crucial but how much rain falls in a particular period is more important. A seasonal perspective was also analysed given that the region has distinct dry and wet seasons.

5.2 Results

In the first part of the results section, we present the analysis based on IMERG observations, starting with the deterministic and then probabilistic verification. The analysis against gauges will be presented at the end of the section.

5.2.1 Deterministic verification

As shown in Fig. 5.1, the study domain has four distinct seasons; December-January-February (DJF), March-April-May (MAM), June-July-August (JJA) and September-October-November (SON). The mean seasonal daily rainfall intensity (dotted-blue contours) show the rainfall maximum in the south in DJF, north in JJA and in the vicinity of the equator in MAM and SON. The highest mean daily rainfall is 8 mm day^{-1} over Lake Victoria in MAM, in the southwestern part of the domain in DJF, and over the Ethiopian highlands in JJA. During each season, some land grid-points, especially in northern part of the domain, are dry, defined here as any grid-point with annual mean daily rainfall of less than 1 mm day^{-1}. These grid-points are masked out in Fig. 5.1 and were excluded in the subsequent analysis. The reforecasts are biased, with overestimation of rainfall over mountainous regions and underestimation in low-lying regions. In JJA when the seasonal rains are concentrated north of the equator, there is an overestimation of $\sim$6 mm day^{-1} over the Ethiopian highlands (Fig. 5.1c) while in the DJF season the overestimation is predominant over the elevated terrain south of the equator accordingly. A feature independent of the season considered is the overestimation over highlands and mountains, e.g., the mountainous Congo-Uganda/Rwanda boarders and Mt. Kilmanjaro in all the seasons (Fig. 5.1). However, when the errors are normalized with their seasonal means, the overestimation in the rainy seasons scale down. Rather, there are large normalized mean absolute errors (NMAE) in the dry seasons (see Fig. A.7), probably due to higher frequency of very low rainfall intensities (>0.2 mm) in IMERG. The accuracy of the reforecast reduces with increase in rainy day threshold (see Fig. 5.1e). In all the seasons, the absolute MAE increases from a mean of 3–4 mm for a rainy day threshold of >0 mm day^{-1}, reaching about 7.5–8 mm day^{-1} at a threshold of >10 mm day^{-1}, depending on the season. The absolute error initially sharply raises from the initial value reaching $\sim$5 mm at a threshold of 1 mm day^{-1}, then continues increasing almost linearly, but at a lower rate. The largest inaccuracy is recorded in JJA, followed by MAM, DJF and SON, based on the domain-averaged MAE and NMAE. These biases can be partially attributed to model deficiencies. However, caution should be exercised here because the observation data set, IMERG, has been known to exhibit a dry bias over high altitudes topography and for high intensity rain rates (O and Kirstetter, 2018; Ageet et al., 2022), possibly accounting for the larger biases.

Despite these positive biases, we see many grid-points with zero or weak bias, indicating low biases at these points. The fact that the model mostly overestimates rainfall during

the rainy season also suggests that it captures the seasonal cycle of the rainfall, putting rainfall in the right locations, albeit with wrong amounts and/or frequency. On the other side, the model underestimates precipitation in some locations, especially over low lands like the Congo Basin in all the seasons. This can be explained by the fact that the model struggles to represent convective rainfall at the mesoscale, often leading to frequent low intensity rainfall (Marsham et al., 2013; Birch et al., 2014; Vogel et al., 2018).

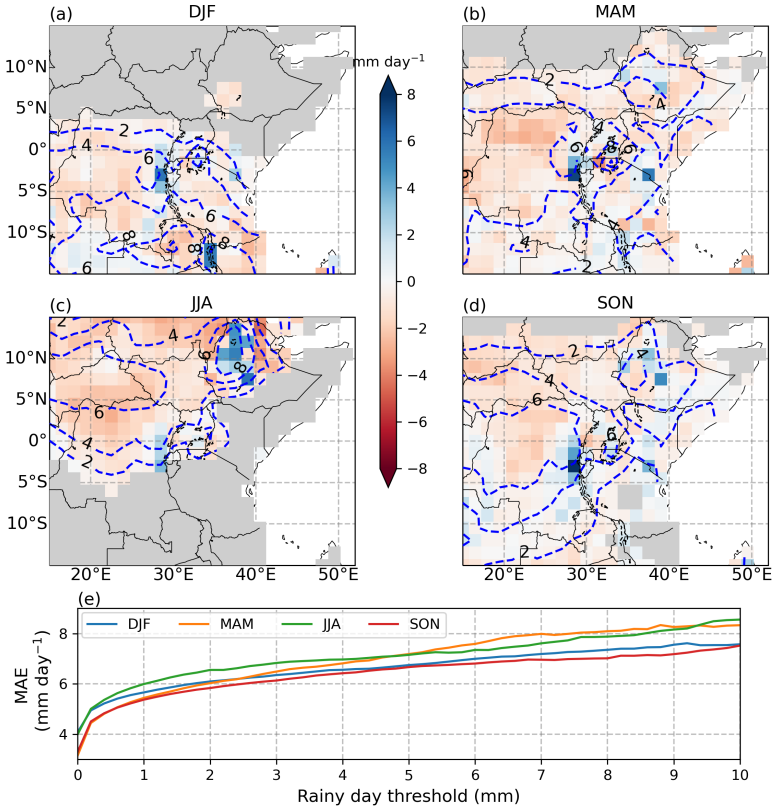


Figure 5.1: Seasonal mean daily rainfall bias of the hindcasts compared to IMERG (a–d). The mean error is computed for only hits, i.e., days when both the hindcasts and IMERG recorded 0.2 mm or more of rainfall. The dotted blue contours represent the daily mean seasonal rainfall computed from the IMERG. Note also that dry grid-points (seasonal mean daily rainfall amounts less than 1 mm day^{-1}) are masked out shown by the grey shading. (e) shows the variation of the mean seasonal absolute error with rainy day threshold. The absolute error is also computed for only hits, i.e., both IMERG and the reforecast have rainfall greater or equal to the threshold. The thresholds used in (e) is 0–10 mm day^{-1} at 0.2 mm intervals. Taken from Ageet et al. (2023), submitted to *Weather and Forecasting*.

5.2.2 Probabilistic verification

5.2.2.1 Raw reforecasts

Over the entire study period, the reforecasts are able to distinguish between rain and non-rain events at 0.2 mm rainy day threshold up to 14 days ahead for some grid-points (Fig. 5.2). The curves in ROC plot are to the left of the diagonal (Fig. 5.2a), with shorter lead-times being the furthest to the left. This implies that the discrimination ability of the hindcasts decreases with lead-time, since the model gradually drifts away from the truth as memory of the initial conditions is eroded with time. EPC15 (black curve) competes favorably with the model, with the equal and higher AUC score after day two and three, respectively. The AUC computed as the area between the diagonal and the curve in Fig. 5.2a, and consequently the AUCS computed relative to EPC15 (Fig. 5.2b–c) confirms the potential skillfulness of the hindcasts in most parts of the study domain. At 1-day lead-time, most of the grid-points have positive AUCS with the strongest values over Uganda, the Ethiopian highland and the southwestern regions. The distribution of the AUCS has an almost southwest–northeast orientation with the better scores over the elevated terrain (cf. Fig. 1.1). Arguably, the model has a better ability to represent orographically triggered precipitation over highlands. Furthermore, but likely of secondary importance, moist static instability over the highlands is not high such that more moisture and higher instability needs to be present, e.g., from larger scale convection signals such as MJO (Pohl and Camberlin, 2006a). This means that over the highlands, the triggers of convection are at larger scales, which are known to have better predictability (e.g., Vitart, 2017; de Andrade et al., 2021; Specq and Batté, 2022), improve the model performance. Over the low-lying regions like the Congo Basin, a low convective inhibition (CIN) and medium convective available potential energy (CAPE) environment, there are more stochastic triggers which the model largely fails to represent (A. R. Satheesh, 2023, in review). The AUCS deteriorates with increase in lead time (Fig. 5.2c,d). The positive AUCS especially at day one has also been previously shown in Walz et al. (2021), although they used the operational forecasts from the earlier version of ECMWF model.

The raw reforecasts have skill in predicting rainfall occurrence relative to EPC15 depending on the time of the year and location (Fig. 5.3a–d and Table 5.1). The hindcasts are able to capture a rainy day in most grid-points in MAM, being up to 40% better than the EPC15. The strongest positive BSS lies along the northeast–southwest stretch in the study domain confirming the potential skill previously seen (cf. Fig. 5.2). In DJF and JJA, most grid-points especially in the locations of maximum seasonal rainfall occurrence have negative BSS values, i.e. the ECMWF model performs worse than EPC15 in predicting a rainy day. This is most likely linked to the fact that the hindcast have a tendency to overestimate the frequency of precipitation during the rainy season (cf. Fig. 5.1). In SON, the model fails to correctly forecast rainy days along

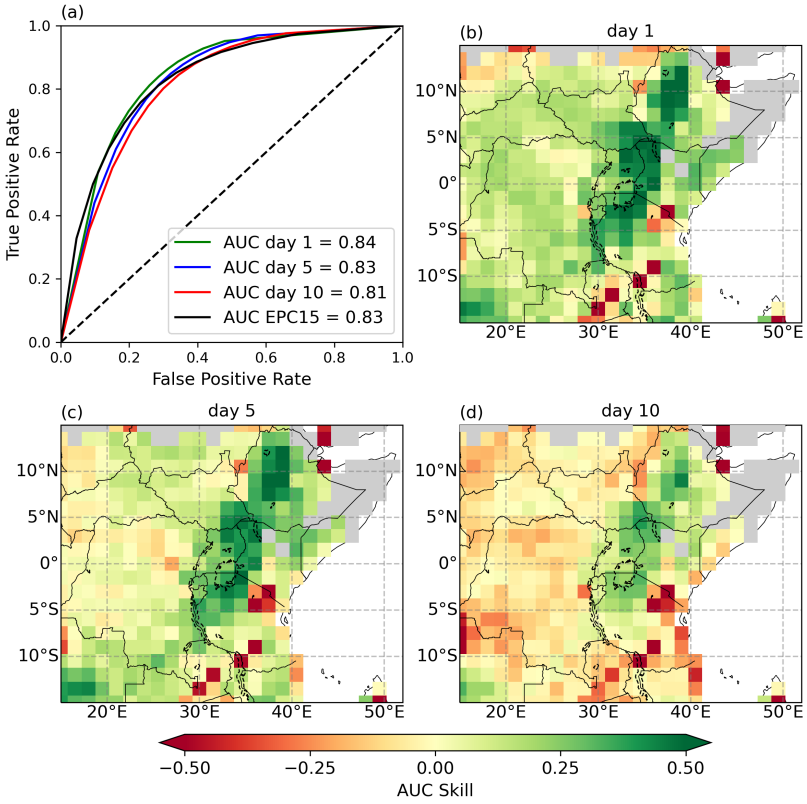


Figure 5.2: Full domain-averaged receiver operating characteristic (ROC) curve (a) for day 1 (green curve), day 5 (blue curve), day 10 (red curve) lead-times and EPC15 (black curve) at a rainy day threshold of 0.2 mm. (b)–(d) are the area under the curve skill (AUCS) computed relative to EPC15. As before, dry grid-points are masked out. Taken from Ageet et al. (2023), submitted to *Weather and Forecasting*.

the East African Coast (EAC) at all the grid-points. Indeed, in all the seasons, the BSS is mostly negative along the coast. This may be due the model's deficiencies in reproducing the sea-land breeze effect, although caution should be exercised here since IMERG struggles retrieving warm rain along the Coast (Vogel et al., 2020). Based on the DM test, the ability of the raw hindcasts to detect rain or no-rain events is significantly different from that of the EPC15 in only 38%, 43%, 31% and 34% of the grid-points in DJF, MAM, JJA and SON, respectively (Table 5.1). Taking the percentage of grid-points with positive BSS values into consideration, we conclude that, although the BS improves for many grid-points relative to EPC15 in all the seasons, with the exception of JJA, the scores are not significantly better in all the grid-points.

The performance of the full distribution of the raw reforecast relative to EPC15 is assessed using CRPSS values shown in Fig. 5.3e–h. The performance of the reforecasts varies spatio-temporally and is generally superior to EPC15, however, the CRPSS are weaker than the BSS values for the same seasons and grid-points. In a few places, e.g., along the Coast, the CRPSS values are higher than the BSS values, similar to the

Table 5.1: Percentage of grid-points where the BSS and CRPSS are positive (Skill) and the score of the reforecasts is significantly different from that of EPC15 (DM) for each of the seasons. The table corresponds to results shown in Fig. 5.3. Taken from Ageet et al. (2023), submitted to *Weather and Forecasting*.

	BSS		CRPSS	
	Skill (%)	DM (%)	Skill (%)	DM (%)
DJF	69	38	74	38
MAM	65	43	70	33
JJA	22	31	33	29
SON	53	34	47	22

general observation in Vogel et al. (2020), who also found that the ECMWF operational model performed better for rainfall amounts. The CRPSS is again best in the DJF and MAM seasons where most of the grid-points having positive values. However, the scores are not significantly different from those of EPC15 in most grid-points (Table 5.1). The worst performance is seen in the JJA season with most of the grid-points having negative CRPSS values. The weaker performance for the CRPSS compared to BSS for the same places, e.g., southwest of the domain and Congo basin in MAM suggest that the reforecast can differentiate between a rainy day, but has biases in the intensities. The opposite is true for some locations like over the western boarder of Tanzania in DJF.

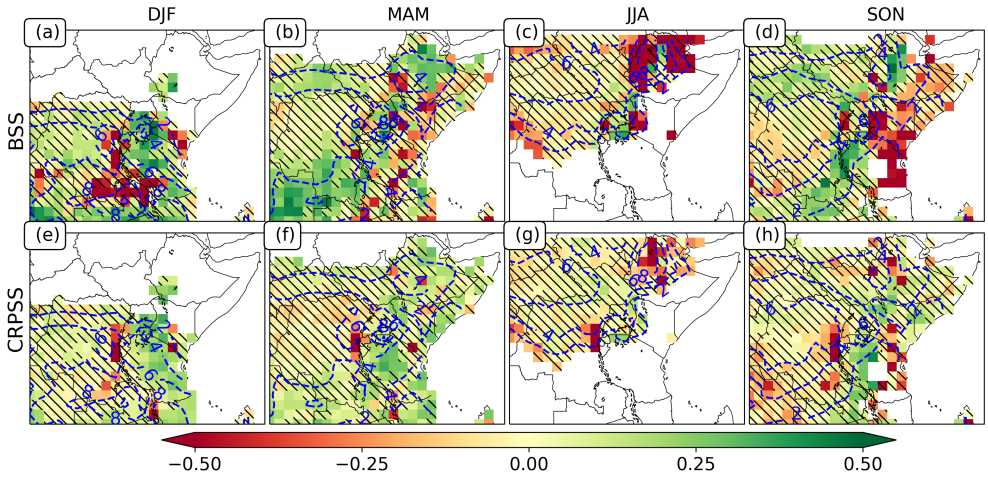


Figure 5.3: BSS and CRPSS of raw hindcast relative to EPC15. The skill scores are calculated for each of the four seasons and only for wet (>1.0 mm day^{-1} in the season) grid-points. The blue dashed contours represent the mean daily rainfall amounts in the season. The hatching shows grids where the hindcasts are not significantly different from EPC15 based on the BS/CRPS values according to the DM test. Taken from Ageet et al. (2023), submitted to *Weather and Forecasting*.

The reforecasts are miscalibrated, as shown in the underdispersed histograms (Fig. 5.4a). The observation frequently rank lowest or highest in comparison to the ensemble, i.e., the observation falls outside of the ensemble range more often than expected. The miscalibration can also be seen in the reliability diagrams for probability forecasts of rainfall occurrence (Fig. 5.4d). The reforecasts are overconfident, that is, when the event happens less (or more) frequently than expected when forecasting high (or low) probabilities. As expected, EPC15 is well calibrated (Fig. 5.4b and e) with the PIT histograms being uniformly distributed and the calibration curve following the diagonal almost perfectly.

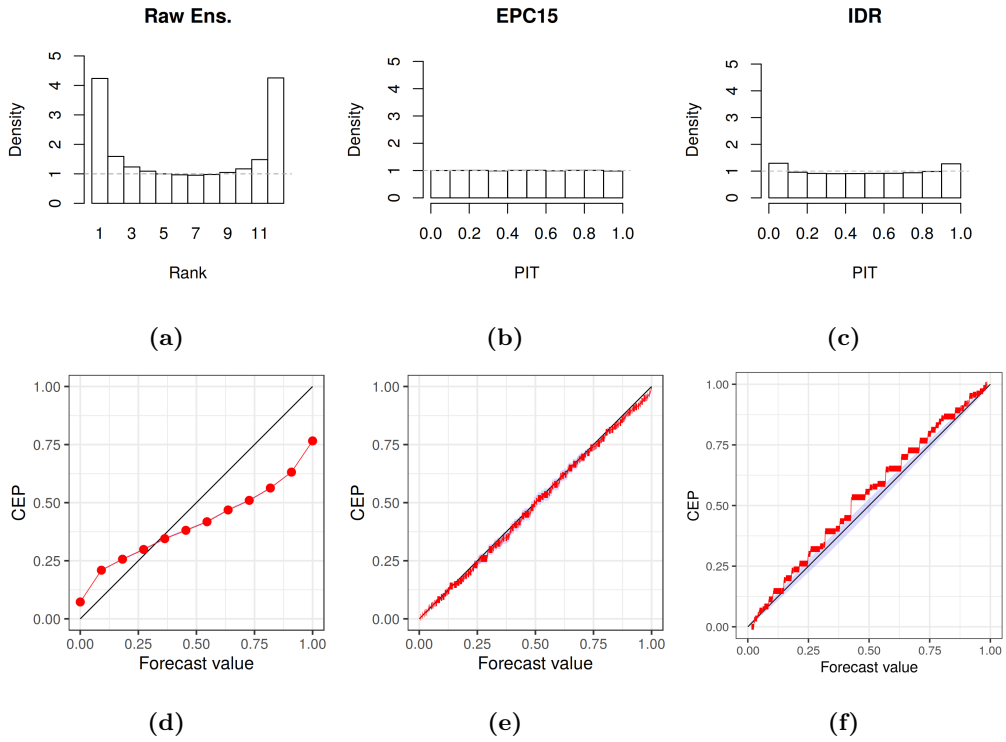


Figure 5.4: Rank histogram at lead-time of one day for the raw ensemble reforecast (a), and the PIT for the reference forecast (b) and IDR-postprocessed reforecast (c) for the period 2001–2018. The lower row plots are the corresponding reliability diagrams. CEP in the y-label stands for conditional event probability. Taken from Ageet et al. (2023), submitted to *Weather and Forecasting*.

Figure 5.5a–c shows the CORP decomposition of the raw BS. The worst MCB is in parts of Eastern Kenya and the around Mt. Elgon area. Mt. Kilimanjaro region and southeastern Ethiopia also have relatively high MCB values. The best DSC is in the southeastern parts of the study domain, stretching from central Tanzania to southern Congo. The model also has relatively high discrimination ability in the northern region, parts of southern Sudan and Ethiopia. The Congo basin features the lowest, i.e. worst, DSC values in the study region, explaining the lower AUCS values discussed earlier.

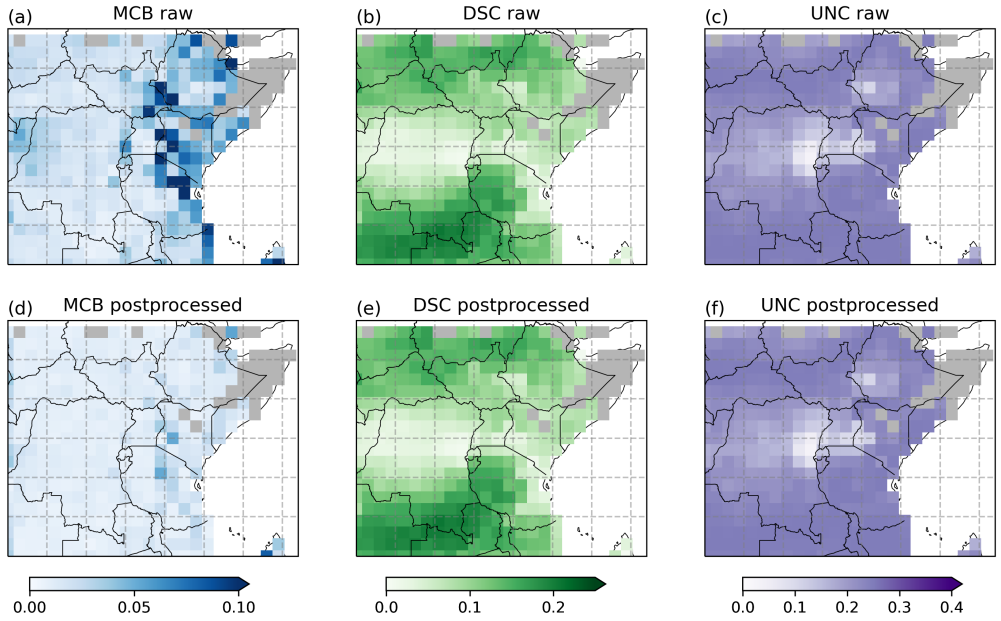


Figure 5.5: Decomposition of the CORP score into the miscalibration (MCB), discrimination (DSC) and uncertainty (UNC) components for raw (a–c) and IDR-postprocessed (d–f) reforecasts. The dry area (grey shading) are masked out as before. Note that the DSC and UNC for raw and postprocessed reforecasts are equal as postprocessing only corrects for miscalibration in a forecast. Taken from Ageet et al. (2023), submitted to *Weather and Forecasting*.

5.2.2.2 Postprocessed reforecasts

Generally, postprocessing using the IDR method reduces the miscalibration. The underdispersion is substantially reduced and the curve is closer to the diagonal in the reliability diagram (Fig. 5.4c,f). Hence, the BSS and CRPSS of the reforecasts improve in almost all the grid-points in all the seasons (Fig. 5.6a–h and Table 5.2). The BSS values are considerably better, especially in MAM, reaching 0.5 in some areas, e.g., the Ethiopian highlands, eastern Uganda and southeastern Congo. The scores are again mostly higher at elevated terrain whereas over the flatter regions, especially the Congo forest basin, they are still low and not significantly different from those of the EPC15. The performance is also better in SON, with only a few areas having high negative BSS values, e.g., over mountainous regions in Uganda (Elgon and Rwenzori), the raised areas in the southeast of Kenya and the mountains over the Rwanda/Burundi-Congo borders. In DJF and JJA seasons, the performance is lower compared to the other two seasons. This is clear at the southeastern Congo and the eastern parts of Tanzania in DJF, where we see high negative BSS values. The same can be seen in JJA over the Ethiopian highlands. A detailed analysis of these two regions revealed frequent cases of 'forecast busts' (almost all ensemble members forecasted rain but it did not rain). These busts resulted in very high BS values for these days, which negatively skewed the BSS

Table 5.2: Same as in Table 5.1 but for the postprocessed reforecasts. The table corresponds to results shown in Fig. 5.6a–h. Taken from Ageet et al. (2023), submitted to *Weather and Forecasting*.

	BSS		CRPSS	
	Skill (%)	DM (%)	Skill (%)	DM (%)
DJF	81	60	91	61
MAM	95	83	84	63
JJA	66	26	64	20
SON	88	56	73	35

in these areas. The reduction in hatched grid-points further highlights the improvement after postprocessing in all the seasons, although to a very small extent in JJA.

The skill with respect to the CRPSS of the postprocessed reforecasts for rainfall amounts also improved (Fig. 5.6e–h). Similar to the BSS, the CRPSS is better in the DJF and MAM and worst in the JJA. Contrary to the BSS in the DJF season, CRPSS around the Southeastern Congo and the eastern boarder of Tanzania is positive. This means that the model is superior in forecasting rainfall amounts than matching rainy days in these parts of the study domain. This feature of the model could be important for some applications like agriculture where the focus is not really on the timing but rather the amount of rainfall in a particular area. In MAM and SON seasons, the lower scores over low-lying areas like the Congo basin region is also apparent. Despite the improvement, the DM test suggests that the performance is not significantly different from that of EPC15 in most grid-points, especially in JJA and SON (Table 5.2). Of note also is that the largest gains in skill after postprocessing is in regions like the EAC which initially have a high degree of miscalibration (cf. Fig. 5.5a, d). For places like the Congo basin where the source of error was the low discrimination ability, the benefit of the postprocessing is almost zero (Fig. 5.6i–p) as DSC was not affected by postprocessing (cf. Fig. 5.5b, e).

Just like in the raw reforecasts, not all the positive BSS and CRPSS are significantly better than the EPC15, given the higher percentages for 'skill' compared to the 'DM' column (Table 5.2). Directly comparing Tables 5.1 and 5.2, there is an average improvement of 82% (67%) in the number of grid-points with positive BSS (CRPSS) after postprocessing. Similar improvements were seen in Vogel et al. (2020) for rainfall occurrence over the tropics. The number of grid-points where the reforecasts are significantly different from EPC15 also increased by an average of 50% and 45% for BSS and CRPSS, respectively. This increase coupled with the positive anomalies in Fig. 5.6i–p, implies that the absolute values of the BSS and CRPSS are also generally improved relative to EPC15.

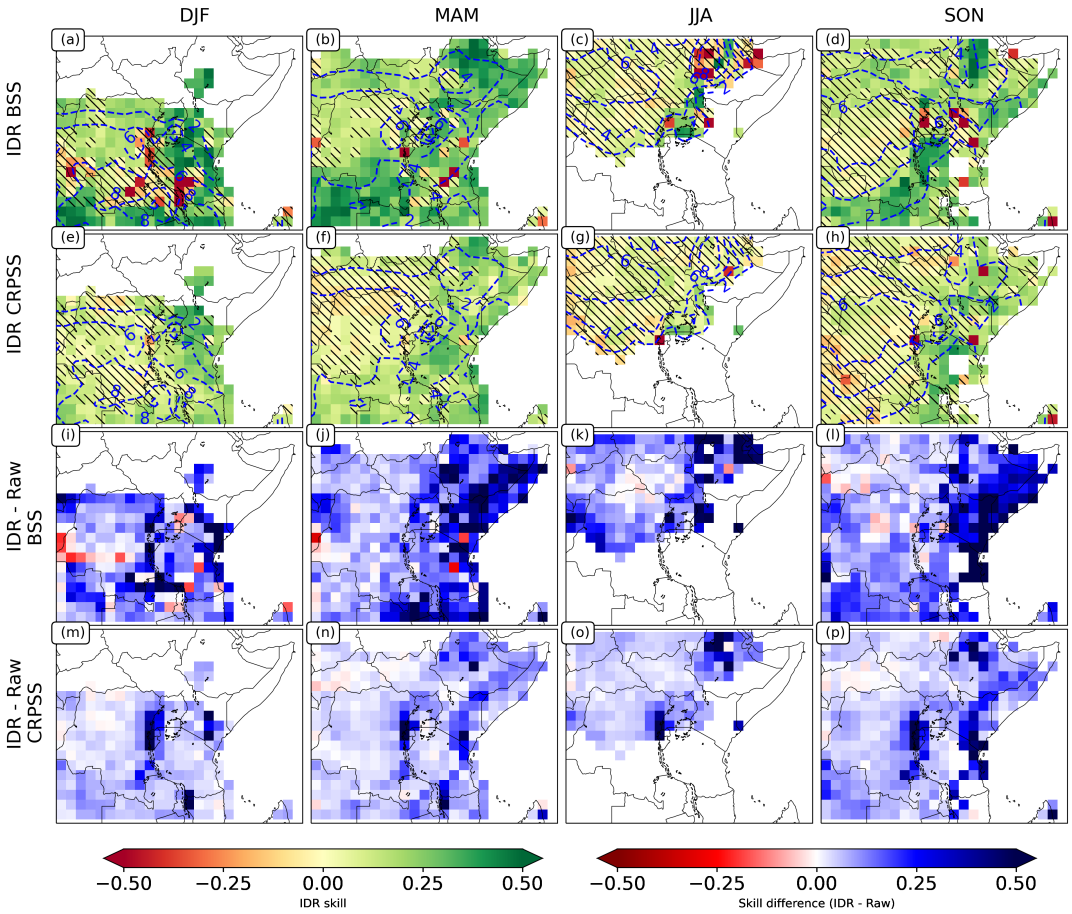


Figure 5.6: (a)–(h) are just like in Fig. 5.3 but for IDR postprocessed reforecasts. (i)–(p) is the difference between the postprocessed and raw reforecast. Taken from Ageet et al. (2023), submitted to *Weather and Forecasting*.

From Table 5.3, the best BS (CRPS) performance is in DJF (JJA) season for the raw and postprocessed forecasts and EPC15. The generally better scores in the dry seasons does not necessarily translate to better skill. This is highlighted by the worst skill scores relative to EPC15 in the JJA season (e.g., cf. Table 5.1). Indeed in JJA, even after postprocessing, the BS of the reforecast is only equal to the score of EPC15, highlighting the limited predictability in this season. Only in DJF and MAM are the raw reforecasts either equal or better than the EPC15, and the scores are further improved after postprocessing. In SON, even though the scores are initially worse than those of the reference forecast, the situation is reversed after postprocessing.

5.2.2.3 Spatio-temporal variation of skill

The skill scores of the ensemble hindcasts vary with location and lead-time (Fig. 5.7). In all the regions (cf. Fig. 1.1), the performance of the reforecasts for occurrence and amounts for 24-hour accumulations degrades with lead-time as expected. Both of the two approaches used here; a) counting the number of grid-points with positive BSS/CRPSS,

Table 5.3: Domain averaged mean seasonal BS and CRPS of raw, IDR-postprocessed corresponding to Fig. 5.3. The numbers in bold (italic) font are the best (worst) scores for each seasons. Taken from Ageet et al. (2023), submitted to *Weather and Forecasting*.

	BS				CRPS (mm day^{-1})			
	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON
Raw	*0.11*	0.15	*0.16*	*0.16*	1.81	2.04	*1.70*	*2.06*
EPC15	*0.11*	*0.16*	**0.12**	0.15	*1.91*	*2.12*	1.56	1.97
IDR	**0.08**	**0.13**	**0.12**	**0.12**	**1.68**	**1.89**	**1.53**	**1.86**

and b) displaying the absolute BSS/CRPSS with boxplots, show this characteristic of the model. We also see the improvement after postprocessing with the skill scores of the postprocessed (red color) being higher than that of the raw reforecasts (black color) in all the regions (Fig. 5.7). In terms of absolute skill score, we find that, although the postprocessing improves the BSS and CRPSS for all the regions, the median scores (yellow line in the box plots) are low positive or negative values, highlighting the poor performance of forecasting systems in the tropics. For some regions, the skill scores are always negative, e.g., EAC for rainfall occurrence detection and Congo basin for rainfall amounts. This worrying level of skill in the tropics has previously been shown by Haiden et al. (2012) and Vogel et al. (2018, 2020).

The best skill scores for both occurrence and amounts of rain are over the Ethiopian highland, followed by the Uganda region. These two regions have relatively higher altitudes, consistent with earlier results which suggested the model performed best over high altitude areas. Over the Congo basin, the results for occurrence, especially for the postprocessed reforecasts is positive up to day five (based on the median scores in the box plots) and over 50% of the grid-points have positive BSS up to day six (Fig. 5.7b). However, for rainfall amounts, the skill scores are always negative after day one (Fig. 5.7f). This region, as explained earlier, is one where the model notably underestimated rainfall and has poor discrimination ability. The performance at the EAC is the worst being negative all the time for the raw reforecats. After postprocessing, this region shows a large improvement in BSS, being positive up to day eight (Fig. 5.7c, g). The fact that raw reforecast in this region had the highest miscalibration explains this high improvement after postprocessing. Similar plots to Fig. 5.7 for higher accumulation of rainfall, i.e., 48-hour and 120-hour (Figs. A.8 and A.9), show a similar trends in the different regions. It is worth noting that with larger aggregation, the BS (CRPS) values get better (worse) as the occurrence mismatches (absolute errors) reduce (grow) (Table S2 in the Suppl. Mat.).

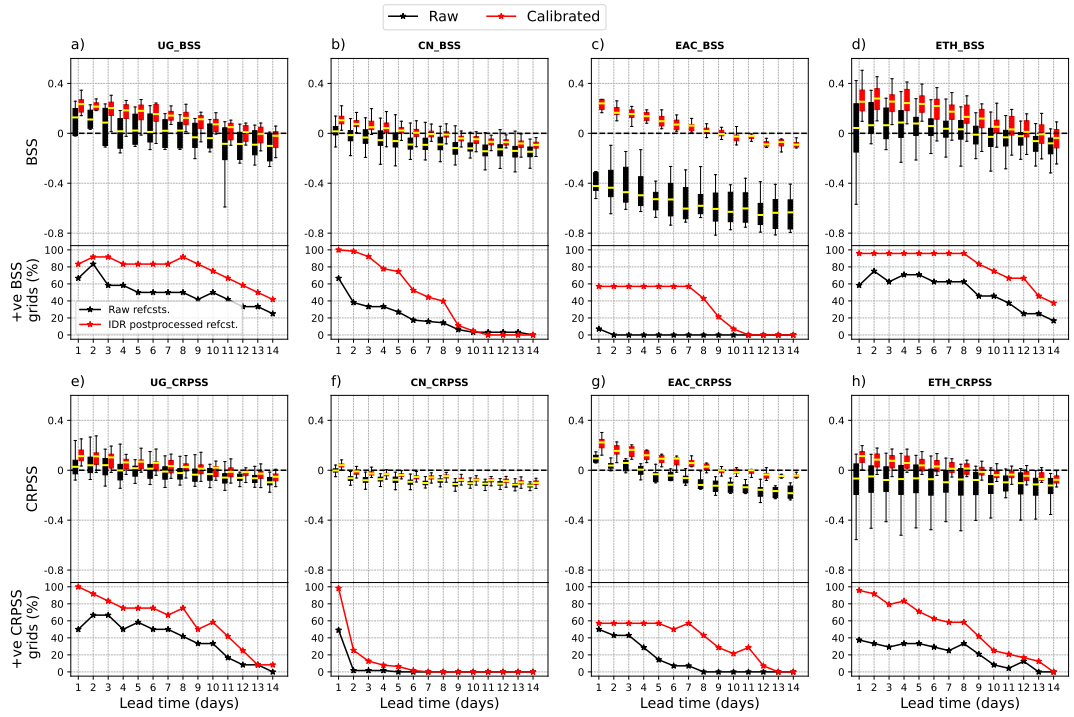


Figure 5.7: Progression of the BSS and CRPSS over lead-time for 24-hour accumulation of rainfall for the four regions (cf. Fig. 1.1) in EEA. The line plots show the percentage of grid-points with positive skill scores relative to all the grids in the particular region, with the black and red being for the raw and postprocessed reforecasts, respectively. The box plots show the distribution of the actual skill scores in the region, with yellow line denoting the median skill score for the raw reforecast with the red line denotes the median skill score of the postpreocessed reforecast. There are 12, 63, 8 and 24 grid-points in the Uganda (UG), Congo basin (CN), East African Coast (EAC) and Ethiopian highlands (ETH), respectively. Taken from Ageet et al. (2023), submitted to *Weather and Forecasting*.

Generally, longer temporal aggregations of precipitation show better BSS values (Fig. 5.8a). This is expected as the error due to time mismatches is gradually reduced. However, it should be noted that this is not the case in all regions, for example over Uganda and Ethiopian highlands, where the skill scores of the 120-hour raw and postprocessed reforecasts rank lowest in performance (Fig. A.10). Similar findings were seen in Vogel et al. (2020), especially in dry areas. They reasoned that 5-day accumulations increased the number of 5-day dry periods in the observation while doing the opposite in the forecasts. We concur with their reasoning that accumulating precipitation for longer time range increases the observation-reforecast mismatches for that accumulation in some regions. Similarly, the model's ability for rainfall amounts also generally improves with increase in the temporal aggregation (Fig. 5.8b).

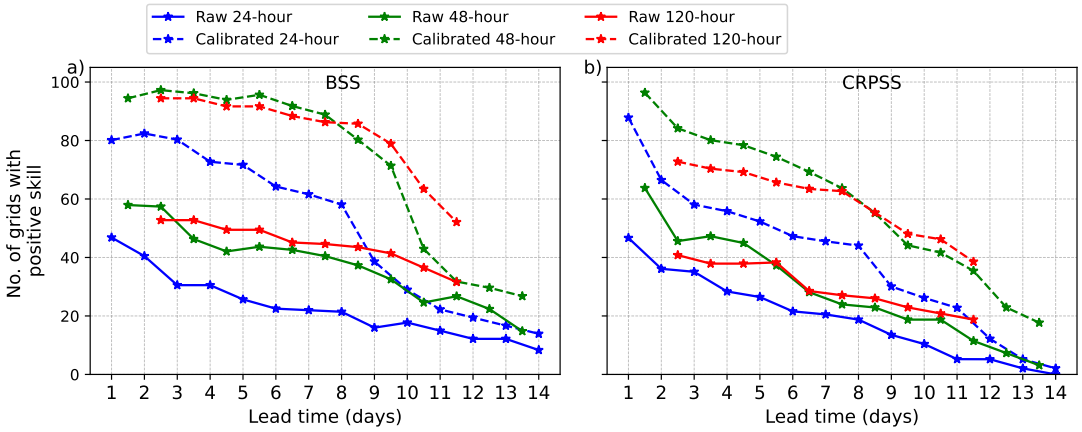


Figure 5.8: Pecentage of grid-points with positive BSS and CRPSS in the raw and postprocessed reforecast for 24-hour, 48-hour and 120-hour accumulation of rainfall averaged over the whole study domain for the period 2001–2018 and increasing lead-time. Note that the x-axis is in "24-hour" interval. For the longer aggregations, the data points lie in the middle of the 24-hour ticks, e.g., day1-2 accumulation data point is located at 1.5 days (the middle point of day1–day2) and day1–5 accumulation data point is at 2.5 days (middle point of day1–day5). Taken from Ageet et al. (2023), submitted to *Weather and Forecasting*.

5.2.3 Assessment based on conventional rain gauges

Because rain gauge measurements are often regarded as the 'truth', we analyse the performance of the reforecast against available gauge data over Uganda. Three grid-points (cf. Fig. 1.1) satisfied the set conditions and the stations used are listed in Table S1 in the supplement. The statistical methods applied are similar to those used in the previous section.

The ensemble forecast more often than not predict occurrence of a rainy day with certainty, i.e., probability of precipitation (PoP) value of 1.0, especially for the Lake Victoria region (Fig. 5.9). Based on the ensemble mean, it rains more in the ensemble reforecast than in IMERG and the gauges (numbers in the top-left of Fig. 5.9). For most days, the observations agree on occurrence of rain (highest frequency of grey shading). However, IMERG has a higher frequency of rainy days compared to the gauges (higher frequency of blue- compared to red-shaded days). This is further highlighted in the reference forecasts; the EPC15 based on IMERG (blue dotted lines) has higher PoP values compared to the EPC15 based on gauges (red dotted lines) in all the three regions. Despite the difference in magnitude of the PoP, both reference forecasts reproduce well the annual seasonal cycle of the rainfall in the different regions.

Postprocessing using IDR modifies the PoP of the forecast, bringing the values towards the EPC15 curves. The IMERG-based postprocessed forecast (green line) shows very little difference though, with most days having PoP value of 1.0, especially during the rainy seasons in the Lake Victoria region. The similarity between the raw and

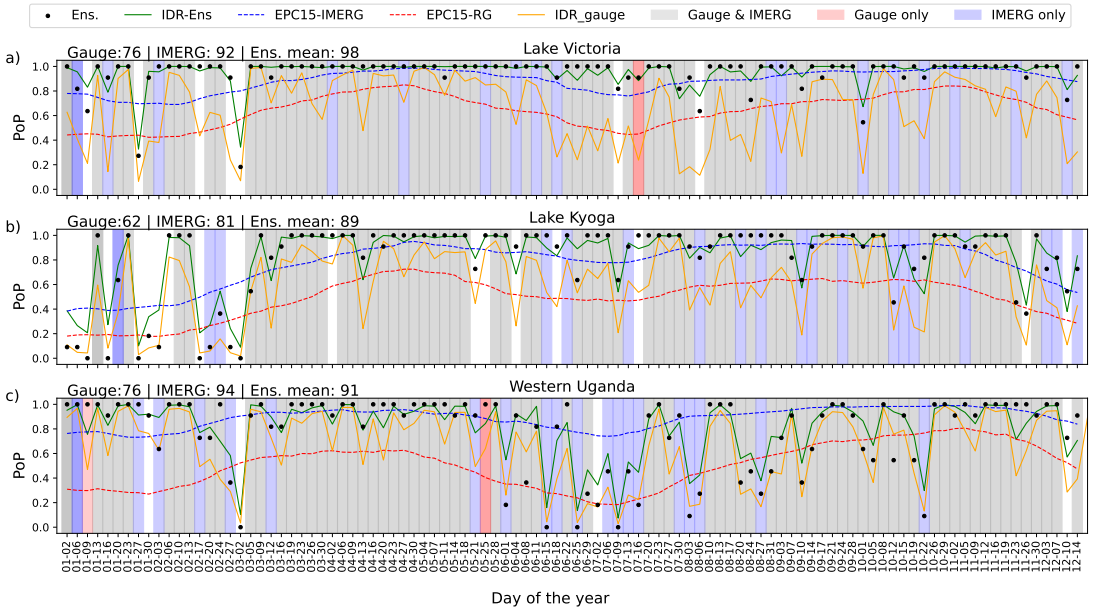


Figure 5.9: Probability of precipitation (PoP) for 24-hour rainfall accumulation evaluated at three grid-points with four gauges (cf. inset of Fig. 1.1) at a rainy-day threshold of 0.2 mm for the different data sets; EPC15 created from IMERG (blue dotted line) and gauges (red dotted line), IDR-postprocessed forecasts using IMERG (green line) and gauge (orange line) observations, and the raw ensemble forecast (black dots). The shading shows when a rainy day was observed in the gauge only (red), IMERG only (blue) and in both gauge and IMERG (grey). The numbers on the top left corner are the counts of the rainy days in the gauge, IMERG and the ensemble mean. Note that only the first 100 out of the 105 reforecast in the year are shown. Taken from Ageet et al. (2023), submitted to *Weather and Forecasting*.

postprocessed forecasts may be due to the fact that, just like in the reforecasts, it rains very frequently in IMERG too, i.e., 87%, 77% and 89% of the time in Lake Victoria, Lake Kyoga and western Uganda regions, respectively, for the year 2001. The gauge-based postprocessing (orange line) modifies the raw forecast to a larger extent compared to IMERG, likely because of the larger differences between the occurrence and amounts of precipitation in the observation and the forecast. We also note that the observation and forecast converge with increase in the number of stations at a grid-point. Note that here we show only the PoP of 2001 as it is representative of the other years (i.e., 2002–2018).

The BSS and CRPSS of the raw and postprocessed forecasts is variable in the three regions, with one common characteristic, that is, the depreciation with lead-time (Fig. 5.10). The best (worst) skill scores are depicted in the Lake Kyogo (Lake Victoria) region for both occurrence and amount of rainfall. The scores are positive, especially the BSS in Lake Kyoga and western Uganda regions, being 35% better than EPC15 at 1-day lead-time. The reforecast perform better for rain occurrence detection compared to accuracy of amounts (i.e., BSS > CRPSS). The CORP decomposition of the BS

shows that the miscalibration of the raw forecasts is worst for the Lake Victoria region and the discrimination ability of the forecast is best in the Lake Kyoga region (Fig. A.11). After postprocessing, the largest improvement in BSS/CRPSS is unsurprisingly seen in the Lake Victoria region because IDR mainly reduces the miscalibration which was largest in this region. In Lake Kyoga and western Uganda, the improvement are smaller due to the initially low miscalibration in these regions.

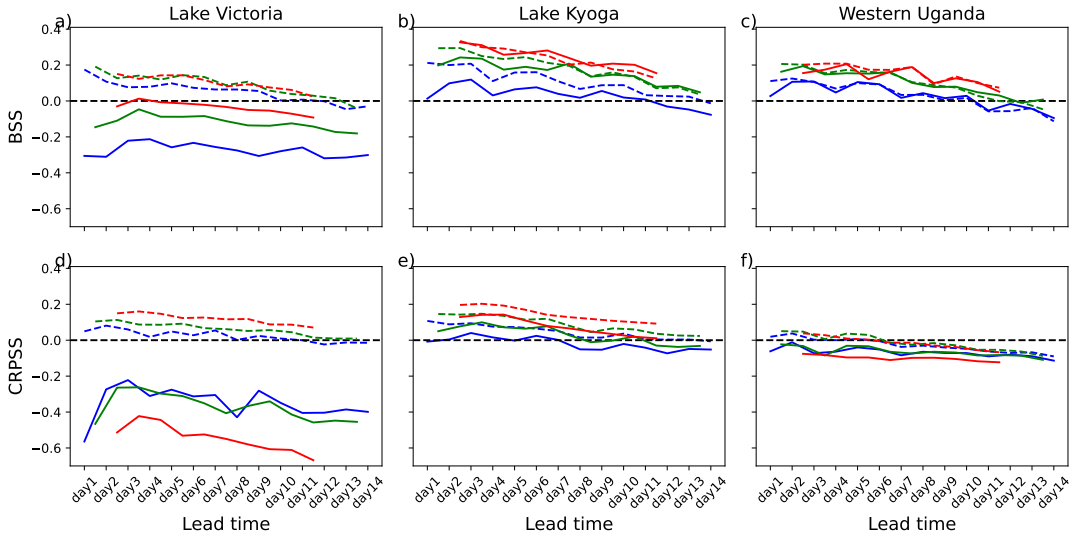


Figure 5.10: Skill scores for raw (solid lines) and IDR postprocessed (dashed lines) reforecasts for 24- (blue line), 48- (green line) and 120-hour (red line) rainfall accumulation in the three regions of Uganda relative to EPC created with gauge precipitation data. Taken from Ageet et al. (2023), submitted to *Weather and Forecasting*.

5.3 Discussions and Conclusions

The study evaluated the skill of rainfall reforecasts from ECMWF against IMERG and gauge observations over EEA for the period 2001–2018. The analysis was done on multiple spatio-temporal aggregations. The reforecasts were analyzed using several verification methods in deterministic and probabilistic terms. The ME and MAE were used to assess the bias and accuracy in the deterministic forecasts, while BS and CRPS assessed the predictive performance of the ensemble reforecast. The reliability diagrams, ROC curves and AUC values were used to asses calibration and discrimination ability of the reforecasts. The skill was assessed using the BSS, CRPSS and AUCS, all computed relative to EPC15. Further, the raw reforecasts were postprocessed to correct the miscalibration, and the resulting forecasts were evaluated using the same verification methods as for the raw reforecasts. The main findings of the study are as follows:

1. The reforecasts are biased, with overestimation of rainfall amounts over mountainous regions. This overestimation is more pronounced during the rainy season. Moreover, the absolute error increases in all seasons with increase in the rainy

day threshold. The overestimation of rainfall observed in the ECMWF hindcasts, especially during the rain seasons and over raised topography agrees with other past studies. de Andrade et al. (2021) and Endris et al. (2021) also showed overestimation over most parts of Africa and Greater Horn of Africa, respectively. They also noted that the overestimation was most pronounced during the rainy season. Overestimation was also shown by Stellingwerf et al. (2021) over Ethiopia, especially for the higher intensity amounts, although they used the ECMWF operational forecasts. As pointed out earlier, the overestimation may not be solely due to model errors since IMERG has also been shown to underestimate warm rain in this region.

2. The raw reforecasts are potentially skillful, being able to discriminate between events and non-events up to day 14, depending on location. This potential skill is translated into positive skill, especially over land with positive BSS and CRPSS, depending on location, lead-time and temporal aggregation. The improvement of the raw reforecasts relative to EPC15 of about 30% (i.e., BSS and CRPSS values of 0.3) in some areas. However, the BS and CRPS values are largely not significantly different from those of the EPC15. The fact that ECMWF reforecasts are skillful confirms results from previous studies which showed that precipitation in this region has higher predictability compared to other regions in Africa and the model is better compared to other models over the region (de Andrade et al., 2021). We also note that the skill is dependent on the season, being best in the DJF and MAM, followed by SON and worst in JJA. This temporal dependence of skill has also been shown by other verification studies in the region (e.g., de Andrade et al., 2021; Endris et al., 2021). We also see that although the skill for rainfall amounts (CRPSS) is highest in the MAM and SON, the magnitude is lower than that of the BSS in corresponding seasons. However, we also note that on average, about 47% (44%) of the grid-points had negative BSS (CRPSS) relative to EPC15. This low or negative skill supports findings in the tropics (Haiden et al., 2012; Vogel et al., 2018, 2020).

3. The reforecasts are subject to biases and calibration errors. Postprocessing using IDR substantially reduces the miscalibration, hence boosting the skill with a 50% improvement relative to EPC15 for most grid-points especially for rainy day occurrence in MAM, DJF and SON seasons. Past studies suggested and showed the benefit of postprocessing (Vogel et al., 2018, 2020; Schulz and Lerch, 2022). We also find that postprocessing the reforecast using IDR considerably improved the BSS and CRPSS over EEA by an average of 82% and 67%, respectively. The largest improvements occur in regions with the highest miscalibration.

4. The analysis against gauges confirms overconfidence of the reforecasts and the improvement after postprocessing. However, it is clear that rainfall in the model

and IMERG occurs more frequently compared to the gauges. Increasing the gauge network helps reduce the bias, pointing to the common problem of point versus gridded data sets comparisons. In our case, we saw an improvement in skill when the number of gauges was increased. We used four gauges in a grid-point which is far from ideal given the 1.5°×1.5° resolution of the reforecasts. However, given the general lack of good and consistent gauge networks in this region, the four gauges in a grid-point is a fair number and provided valuable insights into the skill of the reforecasts.

5. The skill of the reforecasts varies spatially in the study domain, supporting findings of Endris et al. (2021) and de Andrade et al. (2021). The analysis at the different regions revealed that the best performance was over the raised areas of Ethiopian highlands and Uganda. This is partly due the model being able to represent orographically triggered rainfall, and the fact that the convection here is often connected to larger scale signals like MJO (e.g; Pohl and Camberlin, 2006a) which have higher predictability (e.g., Vitart, 2017; de Andrade et al., 2021; Specq and Batté, 2022). The lowest skill scores were over the East African Coast, mainly due to the poor calibration. Over Congo basin, the skill was also poor, owing to the low discrimination ability of the model. The low skill at low-lying regions like the Congo basin has been seen been attributed to the model's struggle to represent convective rainfall (Marsham et al., 2013; Birch et al., 2014; Vogel et al., 2018). The rainfall triggers are stochastic (A. R. Satheesh (2023), in review), limiting the predictability. We acknowledge that although the poor skill especially at the coast is partly due to the model struggles, it has been suggested that IMERG, the observation struggle with warm rain retrieval at the Coast (Vogel et al., 2020), and over high altitudes like mountains over East Africa (e.g., Diem et al., 2014; Monsieurs et al., 2018; Ageet et al., 2022). The analysis against gauges over the Uganda domain further highlight how variable the performance of forecasts can be even over small domains. This apparent variation in skill emphasizes the need for validation studies to ascertain how the model performs in specific regions and not generalize.

This study has highlighted that raw reforecast have skill especially over high altitude areas which is potentially beneficial to meteorological services in the region. However, because the forecasts are biased and uncalibrated, postprocessing is necessary if the forecasts are to offer more meaningful information. Here we used only one reforecast from the ECMWF centre as this has been shown to be the one of, if not the best, in the region. However, studies have suggested that using multi-forecast ensemble mean provides the best outcome (Stellingwerf et al., 2021). This would have the benefit of increasing the ensemble size, especially for reforecasts which often have a limited number of members (e.g., only 11 in our case), hence, increase the spread of the reforecast. We

also tested one novel postprocessing approach, other classical approaches like ensemble model output statistics (EMOS) or Bayesian model averaging can also be applied. Additionally, machine learning approaches have the potential to further improve the quality of the postprocessed forecasts with the ability to incorporate more information than ensemble forecasts of precipitation (Schulz and Lerch, 2022). Given the coarse resolution of the reforecast, it would also be interesting to see how the skill compares if more stations than the four stations used here are included in a grid-point. Studies like Macleod et al. (2021), de Andrade et al. (2021) and Specq and Batté (2022) have shown that skill is regime-dependant in this region. As a next step, we intend to stratify the skill shown here based on known sources of predictability in the region, namely, MJO, IOD and Kelvin waves.

6. Influence of climate drivers on forecast skill

6.1 Overview

Based on the results of verification of the skill of the ECMWF reforecast in the previous chapter, it is clear that the performance of the forecasts varies spatio-temporally. It is also clear that the EEA region's weather and climate may be modified by a range of climate drivers. As suggested in the literature review chapter, these drivers account for the variation in both amounts and frequency of rainfall, and likely the skill of the forecasts too. It is therefore tempting to investigate the role the drivers play in rainfall modulation. More importantly, for a forecasting perspective, the stratification of skill based on prevailing weather and climate drivers may indicate sources of predictability which may be utilisable by forecast users.
In this chapter the influence of climate drivers on the rainfall received at different regions of EEA, and consequently on the forecast skill seen in the previous chapter is analysed. This chapter is meant as a preliminary, but not exhaustive investigation of the role of climate drivers on variability in the rainfall amount received (based on observations) and skill scores of the reforecasts computed in the previous chapter.

The methods used will be introduced in the next section, then the results will be presented and explained in section 6.3. Finally, conclusions and discussion section wraps up the chapter.

6.2 Methods

6.2.1 Wave Filtering

Because EWs propagate zonally in space and time, they can be tracked along the longitude-time spectrum. The propagation is often displayed in a Hovmoeller plot. In this work, four waves types; MJO, Kelvin, MRG and ER have been filtered using the wavenumber-frequency (i.e., space-time 2D domain) filtering technique proposed by Takayabu (1994) and Wheeler and Kiladis (1999). The steps followed are summarised as follows; first, the field to be filtered, IMERG in this case is de-seasonalised to remove the first three harmonics and avoid aliasing (Wheeler and Kiladis, 1999; Knippertz et al.,

2022), the symmetric and antisymmetric components are then computed by adding and subtracting the signals in the corresponding latitudes in the two hemispheres and divided by two, next, the 18-years field (de-seasonalised IMERG) is broken into 96-days segments that overlap by two months. For each segment, the mean and linear trend are removed and the data tapered by zeroes to avoid signal leakage. The segmentation above in turn avoids data loss due to tapering. In the next step, complex FFTs are performed in longitude for each latitude and each time, followed by further FFTs to obtain $k - \omega$ for each latitude. The $k - \omega$ are averaged for all the segments before summing over all the latitudes gives a smooth background spectrum. To obtain the peak power of the individual waves (cf. Fig. 2.8), an inverse FFT is performed taking into consideration only the Fourier coefficients of the specific waves (see Table 2.1) while setting the other windows to zero. Details of the method can be found in Wheeler and Kiladis (1999).

6.2.2 Local phase-amplitude diagrams and creating composites

In order to characterised the waves at each location and time, indices are developed, such as the commonly used global RMM index of Wheeler and Hendon (2004). Although this method has the advantages of showing signal propagation of the MJO with minimal noise (Schlueter et al., 019b), they may not precisely depict the behavior of the signal at specific locations. Therefore, local phase-amplitude indices are generated for the different longitude points of interest using the method introduced in Riley et al. (2011) and has since been applied in many other studies (e.g., van der Linden et al., 2016; Schlueter et al., 019b). The filtered fields, which are usually normalised, are plotted against their local time derivatives. Based on the phase angle and amplitude on a given day, composites of data are constructed. The composites are constructed for each of the eight defined phases, with phase 5 (phase 1) having positive (negative) anomalies, hence they are termed wet (dry) when the filtered field is precipitation (IMERG in this case). Phases three and seven are neutral, while the rest are transitional phases. A phase is said to be inactive when the amplitude is less than one standard deviation (Schlueter et al., 019b). Plotting a scatter displays the transition of the wave in a circle across the different phases over a period of time. The track followed is especially smooth and clear for the slow moving signals like the MJO.

6.2.3 Rainfall modulation

To assess the modulation of rainfall by the EWs, composites of rainfall data is created based of the local phases described in the previous section. Only dates when the phase is active (amplitude >1) are considered. After creating the composites of IMERG data, the data is further split into the different seasons (i.e., DJF, MAM, JJA and SON). The seasonal anomalies are plotted for the whole domain to show which areas have enhanced or reduced rainfall for a given EW during a specific phase. The mean daily

rainfall anomalies for each of the four regions demarcated in Fig. 1.1, i.e., Uganda (UG), Congo Basin (CN), East African Coast (EAC) and the Ethiopian highlands (ETH) are computed in order to quantify the modulation.

6.2.4 Regime-dependent skill

After assessing the modulation of rainfall by the EWs, their impact on the capability of ECMWF reforecast to predict rainfall over EEA is analysed. To ensure that results are robust, only waves that strongly modulate rainfall are considered for this analysis. Secondly, the analysis is done for locations where the modulation is strongest for a given wave. Just like in the previous section, composites of the skill score (BSS and CRPSS) in different active phases of the waves are analysed. The skill scores used are those computed in chapter five.

6.2.5 Role of the Indian Ocean Dipole

In a similar fashion to the waves, the role of the IOD in modulations of rainfall over EEA is briefly analysed. Moreover, since IOD and ENSO have a large co-variability, with the former contributing most of the variability during fall an early winter (IPCC, 2021). Therefore, by assessing IOD, ENSO is implicitly assessed. The daily rainfall accumulation and reforecast skill scores are stratified into days of positive, neutral and negative IOD index. Because IOD predominantly affects the short rains in SON over East Africa (e.g., Black et al., 2003), only this season is considered for this analysis. The IOD indices are taken from NOAA website (https://stateoftheocean.osmc.noaa.gov/sur/ind/dmi.php). During the period considered in this work (i.e., 2001–2018), the negative phase of the IOD happened in only two years (2010 and 2016). Therefore, the analysis was done for two SOND seasons, i.e., two years (2006 and 2015) with positive, two years (2009 and 2008) of neutral and the two years of negative IOD indices.

6.3 Results

6.3.1 Wave filtering

Figure 6.1 shows the propagation of the waves in the MAM season of 2002 based on filtered IMERG daily rainfall for the latitude band 15°S and 15°N. This year and season were chosen because the MJO amplitude over EEA was high in 2002, and MAM is the most important rainy season across most of EEA. The MJO and Kelvin wave propagate eastwards, while the MRG and ER propagate westwards (not shown). The MJO is the slower moving signal at a speed of about 7 ms^{-1}, while the Kelvin wave is faster, propagating at about 12 ms^{-1}. Both the MJO and Kelvin wave modulate rainfall quite strongly, with the anomalies ranging between -10–12 mm d^{-1} and -15–18 mm d^{-1} for MJO and Kelvin, respectively. These characteristics of the individual waves seen here agree with what is known from theory (e.g., Matsuno, 1966; Takayabu and Nitta, 1993;

Wheeler and Kiladis, 1999) and what has been shown in multiple studies of MJO and EWs activity (e.g., Schlueter et al., 019b; Knippertz et al., 2022; Yang et al., 2021).

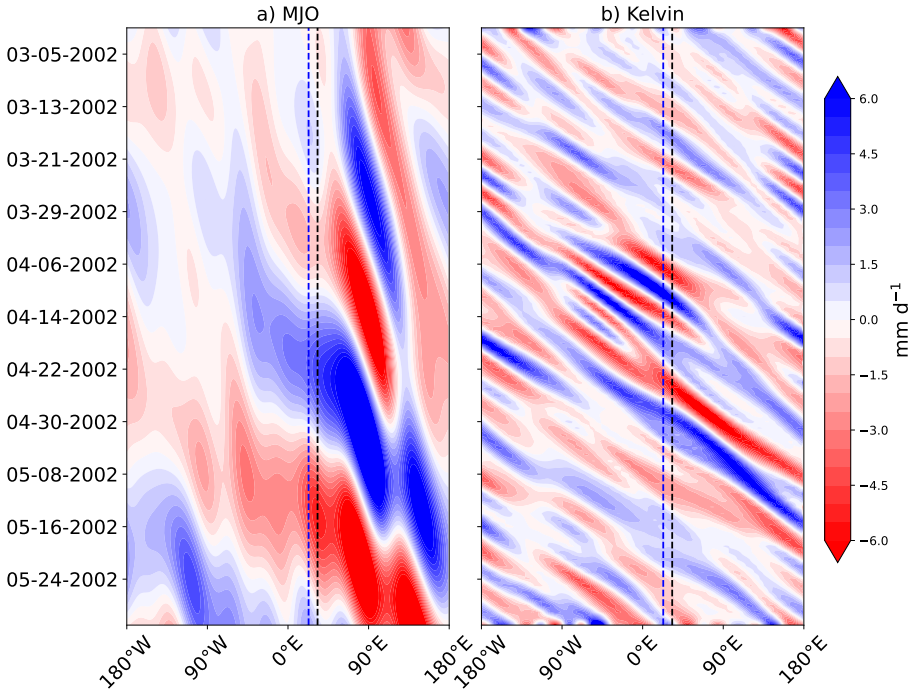
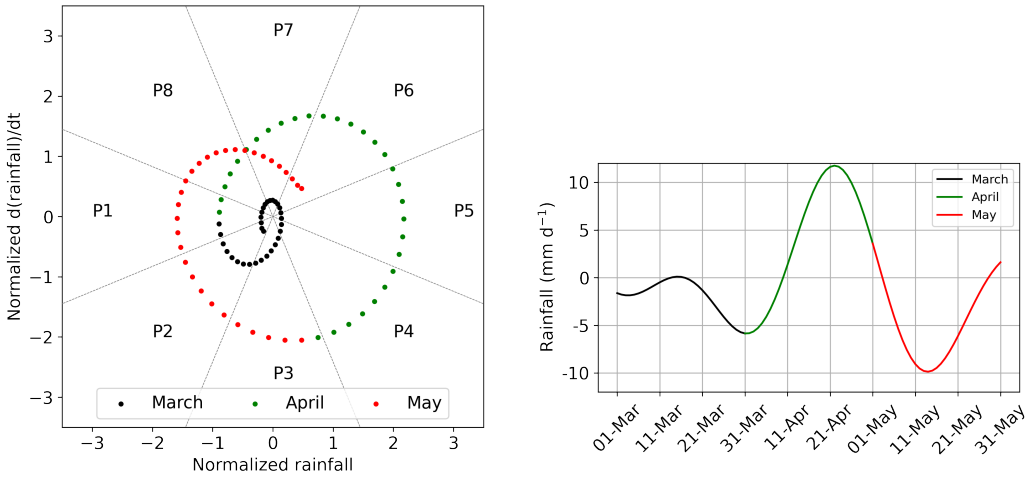


Figure 6.1: Hovmoeller plots of IMERG daily rainfall filtered for MJO (a), and Kelvin wave (b) during the MAM season of 2002. The filtering is done using the wavenumber-frequency approach of Wheeler and Kiladis (1999). The anomalies are computed against the de-seasonalized 2001–2018 climatology. The blue (red) shading imply increased(suppressed) rainfall at the specific longitude and date. The blue and black dotted lines drawn at 23°E and 33°E are points over EEA and where the local phase-amplitude diagram are based. These two points represent Uganda and Congo, respectively.

Figure 6.2(a) shows the propagation of the MJO through the eight phases defined at longitude 33°E. During the MAM season of 2002, the MJO was mainly active, with its amplitude being greater than one standard deviation most of the days. The wave is mostly in the positive (negative) phase in April (May). The impact of the wave on rainfall amount received can be seen in Fig. 6.2b, whereby there is a positive (negative) anomaly in the order of 10 mm d^{-1} during phase 5 (phase 1) which is the wet (dry) phase (Riley et al., 2011; Yasunaga and Mapes, 2012; van der Linden et al., 2016; Schlueter et al., 019b). Although studies like Berhane and Zaitchik (2014) and Vashisht and Zaitchik (2022) suggested that the impact of MJO is suppressed during April, this was contrary in 2002 with the highest modulation seen in this month. One possible explanation for this discrepancy could be the fact that these studies used the global RMM indices (Wheeler and Hendon, 2004). But could also be simply due to the fact that MJO varies monthly but also annually (Suhas and Goswami, 2010). Similar plots can be made for the Kelvin wave. However, because it is a fast moving signal, the points

are random and do not smoothly trace a track for daily data as in the case of MJO. With shorter temporal accumulations, e.g., 3-hourly or 6-hourly, the Kelvin wave track is slightly smoother as can be seen in Yang et al. (2021) (their Fig. 10).



(a) **(b)**

Figure 6.2: a) Local MJO phase diagram for filtered IMERG rainfall at 33°E (black dotted line in Fig. 6.1) against the time derivative at 33°E. The time period considered is the same, MAM season of 2002. The corresponding modulation of rainfall at this point during the period is shown in (b). Phase 5 (mostly occurred in April) is wet while phase 1 (mainly in May) is dry.

6.3.2 Rainfall modulation by equatorial waves

Following the phases identified in the previous section, composites of IMERG rainfall were constructed to analyse how the waves modulate rainfall amounts in the region. Because of its importance, mainly the results of the MAM season are shown. Results for the other seasons are shown where necessary.

Figure 6.3 shows the rainfall anomalies in MAM for composites of the dates falling in the different MJO phases defined at 33°E (the black line in Fig. 6.1). The composites are based on the whole 2001–2018 period. Note that only dates when the MJO is active are considered, giving between 113 and 160 days in a single phase. It can be seen that the rainfall anomaly varies quite a lot with location and phase. During phases 1–3, the region is largely drier than normal while in phases 4–6 most areas, especially in the eastern part of the study domain are wetter than normal. The peak positive (negative) anomaly is in phase 2 (phase 5). The anomalies vary from more about -4 to 6 mm d^{-1} based on regional daily averages (numbers in bottom left corner of Fig. 6.3). During the wet phase, the highest modulation is over the East African Highlands, consistent with finding of Pohl and Camberlin (2006a), Pohl and Camberlin (2006b), Berhane and Zaitchik (2014), and Vellinga and Milton (2018). These previous studies also explain

that during the wet phase of the MJO, moisture from the Indian Ocean is advected towards East Africa leading to enhanced convection. The propagation of the MJO eastward can be inferred from phase 3 through to phase 7 (Fig. 6.3).

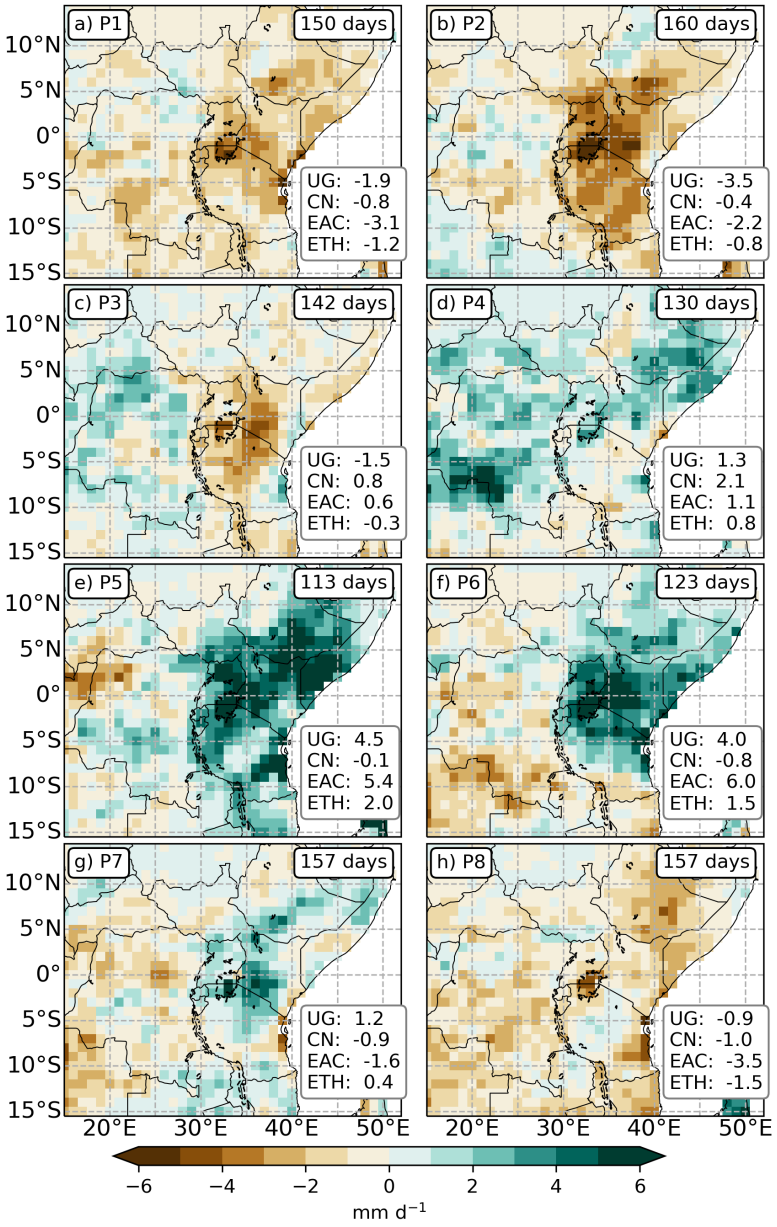


Figure 6.3: Rainfall anomalies of MAM rainfall in the eight phases of MJO defined at 33°E. The numbers at the top right corner are number of days available in a given phase indicated at the corresponding top-left corner. The numbers at the bottom left are region-averaged daily rainfall anomalies.

Similarly, Figure 6.4 also shows rainfall modulation but by the Kelvin wave. The local phases are determined at 23°E (the blue line in Fig. 6.1), which represents Congo, for the same period, i.e., MAM season of 2002. Similar to MJO, each phase has between 124 and 150 days. Phase 5 is the wet while phase 1 is dry. The fact that Kelvin wave modulates rainfall supports findings of Mekonnen and Thorncroft (2016) and Ayesiga

et al. (2021). The later noted that the Kelvin wave modulates rainfall by as much as 5 mm d^{-1}. These values are comparable to the results achieved here (maximum of 7 mm d^{-1} over Congo). One stand out characteristic of the Kelvin wave over the region is that its impacts on rainfall variability is very strong and localized. By eye and comparing the numbers in the average regional anomalies (numbers in the lower-left corner), the difference in modulation over the Congo compared to other regions, and the MJO (cf. Fig. 6.3) are large. Ayesiga et al. (2022) showed that the Kelvin wave has one of its peak activity at 10°-20°E, longitudes were Congo partly lies, the second peak being over the Indian Ocean. Since the Kelvin wave is more constrained to the equator compared to the other EWs, it is not very important at off-equator locations like ETH.

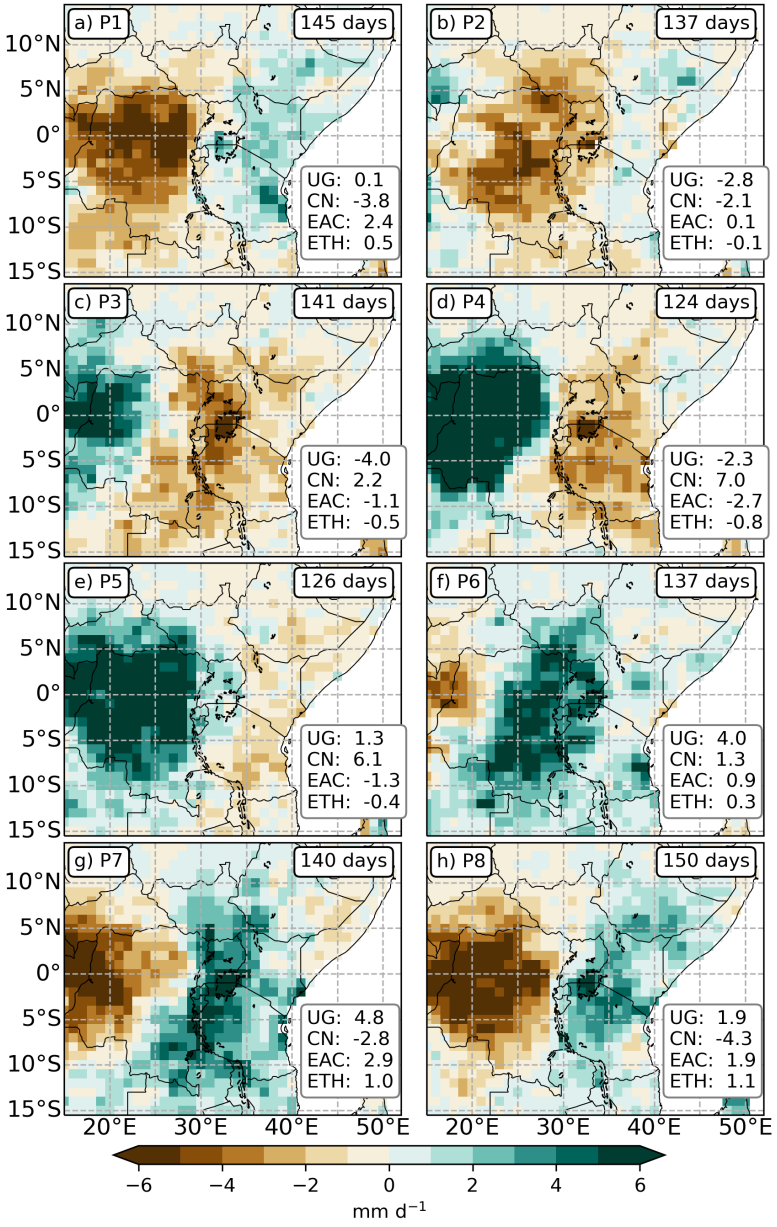


Figure 6.4: Same as Fig. 6.3 above, but for Kelvin wave phases defined at 23°E

Rainfall modulation by the MJO and Kelvin waves happens in other rainfall seasons too (Fig. 6.5). Clearly the waves are dominant in different locations; the Kelvin wave being stronger over Congo basin while the MJO over Uganda and the EAC.
In MAM, the MJO affects rainfall in all regions, with the smallest rate of modulation being over the Ethiopian highlands (red line). This is perhaps not surprising given that MJO is also centred at the equator where its strongest impacts are felt. The peak modulation is generally in phase 5 (also seen in the Fig. 6.3) when daily rainfall accumulation deviate by as much as 5 mm d^{-1}. This variation concurs with what has been reported in previous studies. Pohl and Camberlin (2006b) estimated that 44% of rainfall variability during the wet season can be attributed to MJO. During the JJA season, MJO seems to have less impact on rainfall. This is consistent with studies of Berhane and Zaitchik (2014) and Vashisht and Zaitchik (2022) which also showed that the MJO is mostly pronounced in March, May, and November–December months.

The Kelvin wave strongly modulate rainfall over Congo (magenta lines) in all seasons in equal measure. This is agreement with Ayesiga et al. (2022) who showed that the Kelvin wave is active all year round. The rainfall increases by ~7 mm d^{-1} during phase 5 and decreases by ~5 mm d^{-1} in phase 1 compared to the climatology. Again, these values are similar to the 5 mm d^{-1} modulation in Ayesiga et al. (2022). The effect of Kelvin waves on rainfall is smallest over the Ethiopian highlands, most likely due to its being further away from the equator. Similarly over the EAC, the effect is small, except in the MAM season when rainfall is reduced by ~2.5 mm d^{-1} in phase 4. Over Uganda, the Kelvin wave modulates rainfall substantially, especially in the wet seasons. The highest modulation is in MAM when rainfall is modulated by as much as 5 mm d^{-1} in either directions depending on the phase (Fig. 6.5f).

6.3.3 Variation in reforecasts skill

Having confirmed that EWs, specifically, the MJO and Kelvin waves modulate rainfall, the next logical step is to assess their influence on the skill of reforecasts. To do this, the skill calculated in the previous section is composited in a similar fashion to IMERG rainfall in the previous section. Given the reforecasts analysed were initialised twice weekly, fewer dates per phase, i.e., between 26 and 42 days per season. Like for the rainfall modulation, only days when the waves are active (>one standard deviation) are considered.

Figure 6.6 shows variation of the BSS amd CRPSS with the phase of MJO and Kelvin wave defined at 33°E longitude. In all the seasons, the wet phases have higher scores for rainfall occurrence (Fig. 6.6a–d) compared to the dry phases. The best skill scores are in phase 5, the phase where the highest positive modulation of rainfall happens. The Kelvin wave does not vary skill of rainfall occurrence in this location (blue line is almost flat). The improved BSS values for MJO are likely due to the increased frequency of

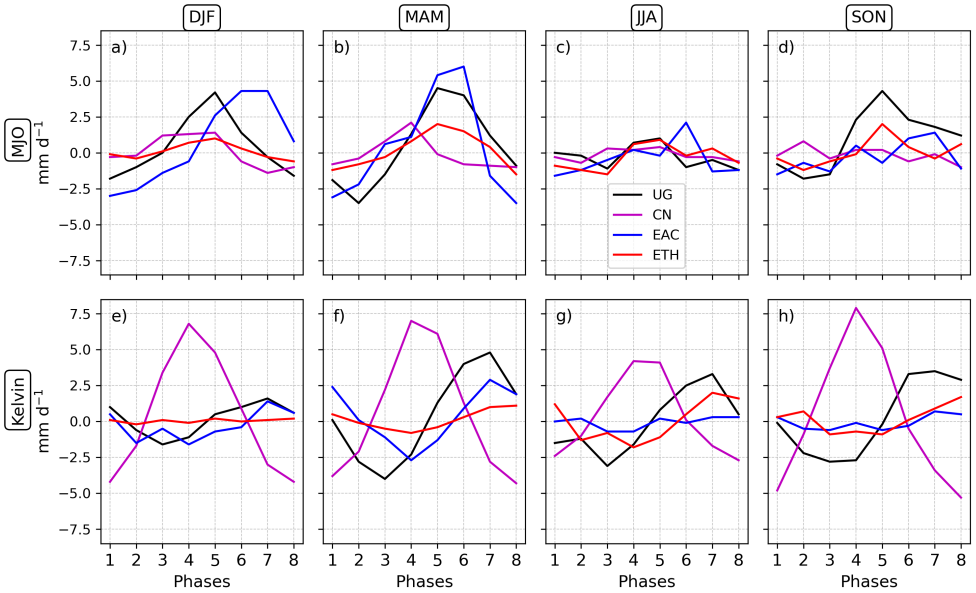


Figure 6.5: Variation of rainfall anomalies averaged over the four regions in EEA during the different phases of MJO and Kelvin wave. Note that the phases are defined at locations where the wave impact is stronger compared to other regions in the study domain, i.e., 33°E and 23°E for MJO and Kelvin wave, respectively.

rainfall occurrence, which in turn increases the hit rate, hence, improved Brier scores. After postprocessing (dotted lines), the skill improves in all the phases with the highest improvement also in the wettest phases, further strengthening the argument made in the previous sentence. Averaging the scores over all the seasons gives a general picture. The predictability of rainfall occurrence is improved over Uganda region when the wet phase of the MJO prevails. On average, the difference in BSS between the wet and dry phase is ~0.2. Similar findings where shown by Specq and Batté (2022) who also found that the MJO improved the predictability of rainfall occurrence. They found that this improvement was due to an increase in the hit rate resulting from a higher rainfall intensity. Also of note is the generally better scores in DJF and JJA seasons, consistent with the results in chapter five of this thesis.

The skill for rainfall amounts assessed using the CRPSS (Fig. 6.6e–h) shows that the MJO does not have an obvious influence on skill. The score is largely the same in all the seasons (almost flat red line in the bottom raw). This could be because most likely the increase in rainfall intensity seen IMERG in Figure 6.5 is matched in the reforecasts. However the Kelvin wave seems to vary the CRPSS. A closer look reveals that the skill is best in phase 4, when the rainfall is negatively modulated by the Kelvin wave over Uganda (cf. Fig. 6.5e–h, the black line). This suggests the reduced rainfall during phase 4 of the Kelvin wave over Uganda means, lower biases, hence higher CRPSS. The skill changes by a maximum of ~0.15 between the dry and wet phases.

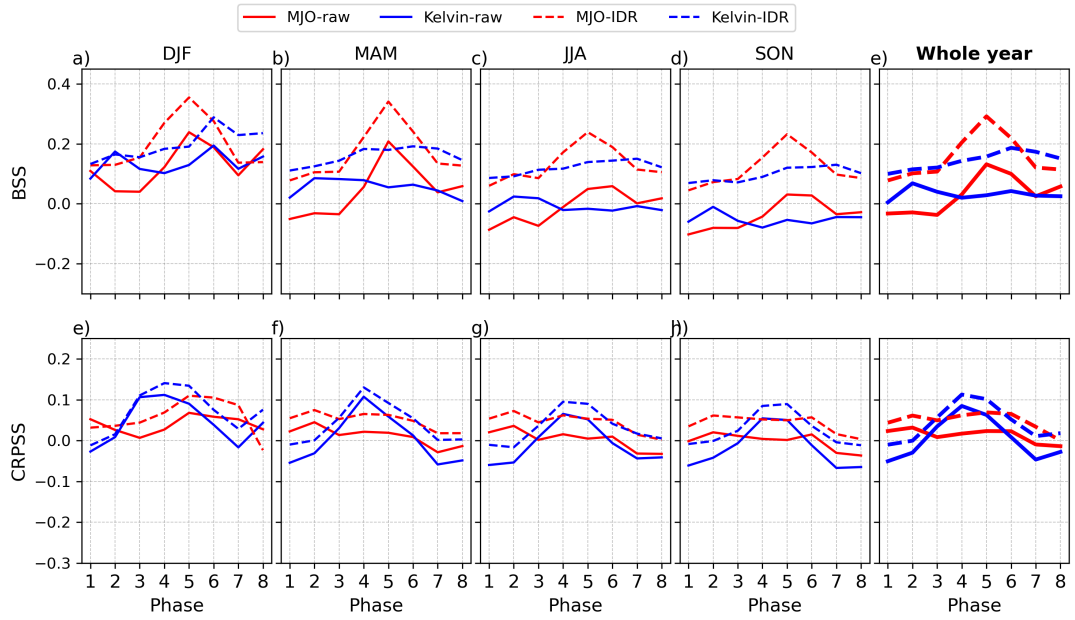


Figure 6.6: Composites of the BSS and CRPSS during the eight phases of MJO (red line) and Kelvin wave (blue) line, with the phases locally determined at the 33°E (black line) in Fig. 6.1. The solid lines are the raw skill and dotted lines for the postprocessed skill scores.

Figure 6.7 shows the influence of the MJO and Kelvin wave on skill scores over Congo area (23°E). The Kelvin wave is more dominant over this region compared to the other EWs and MJO, confirming the results in section 6.3.2. For rainfall occurrence, the BSS (Fig. 6.7a–d) also is better in phases where rainfall is enhanced. The maximum positive rainfall anomaly and skill, both occur in phase 4. This strengthens the results seen in Figure 6.6, and the suggestion that enhanced rainfall means more frequent rainfall in both the reforecasts and observation that result into heightened hit rate. One striking observation here is the fact that contrary to expectation, postprocessing does not improve the score during the wet phase. Score for the postprocessed reforecasts are better during the dry phase, although the anomaly is almost zero. A possible explanation could be that over the Congo basin, miscalibration was minimal (cf. Fig. 5.5). Because IDR only corrects for miscalibration of the forecast, postprocessing does not improve the reforecasts in this region. The better skill in the raw reforecast could possibly be because the postprocessing process included eight surrounding grids to train the model at specific grids, which may have introduced error instead.

The CRPSS pattern is similar to that of the BSS for the Kelvin wave at the Congo location too (Fig. 6.6). Again, the MJO does not modulate the CRPSS, supporting the early suggestion that the influence of MJO over the Congo basin is less pronounced compared to that of the Kelvin wave.

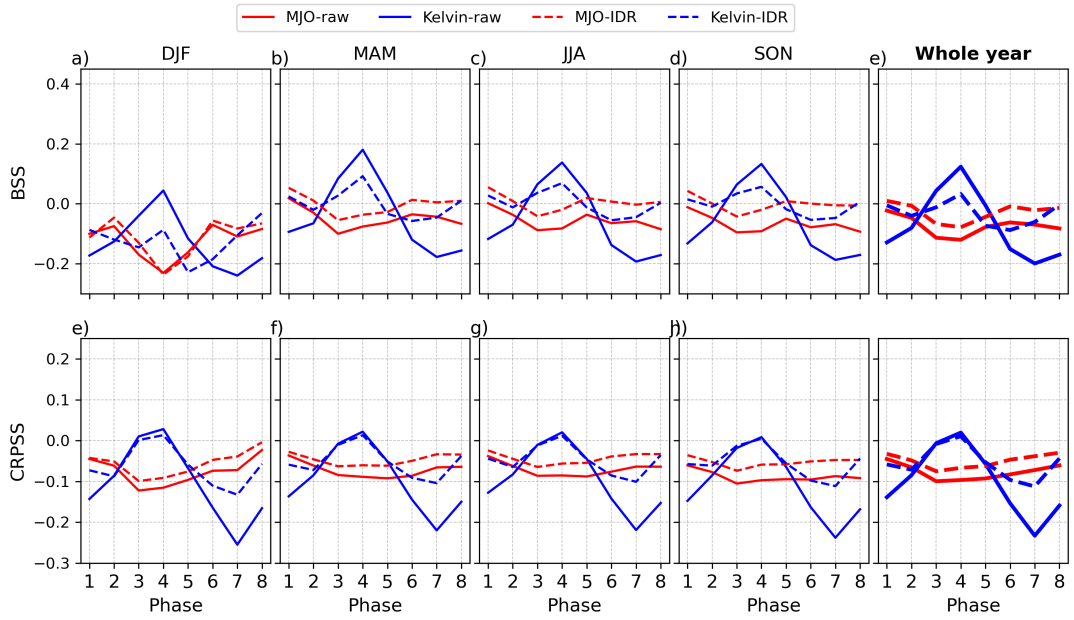


Figure 6.7: Same as Fig. 6.6 but for the phases drawn at 23°E (blue line in Fig. 6.1.

6.3.4 Impact of Indian Ocean Dipole on Reforecasts skill

As explained in Chapter two, rainfall varies at interannual time scales too. IOD has been cited as one of the variability modes responsible. Figure 6.8 shows rainfall anomalies and reforecasts skill in the three phases of the IOD. Only the short rainy season (defined as September–December, SOND, for this analysis) since it is the one most impacted by IOD (e.g., Black et al., 2003; Hirons and Turner, 2018).

The IOD modulated rainfall in EEA (first column), especially over UG and the EAC (Figs. 6.8a,g). During the positive phase (blue shading), there is a positive anomaly in the SOND rainfall while during the negative IOD phase, the rainfall is reduced. In the neutral phase, the rainfall anomaly is in between that of the positive and negative phases. Similar findings have previously shown the role of IOD in rainfall modulations (e.g., Black et al., 2003; Hirons and Turner, 2018; Wainwright et al., 2021; de Andrade et al., 2021). For example de Andrade et al. (2021) found that rainfall in the SON season is correlated to IOD, while Wainwright et al. (2021) show that the extreme rainfall that caused floods over East Africa between October 2019 and January 2020 were as a result of a strong IOD event. Over the Congo basin, the impact of IOD is very negligible. This could be because this region is relatively far away from the Indian Ocean, hence, the moisture advection inland does not reach the region. Over Ethiopia, the IOD may also be relatively insignificant given that median of the rainfall anomalies in the three phase lie almost perfectly along zero (Fig. 6.8j).

The second and third columns show the BSS and CRPSS composites for the three IOD phases, respectively. The skill scores are for the raw reforecats that were computed in the previous chapter. The BSS values in the two regions where the rainfall was visibly modulated, i.e., UG and EAC better during the positive phase. Just like for the cases of Kelvin wave and MJO, the enhanced frequency most likely leads to a higher hit rate during this phase, hence, enhancing the BSS. On the other hand, the skill is worse for the negative phase in both locations. de Andrade et al. (2021) found that IOD improves reforecasts associations over EEA during the SON season. However, they also note that when the influence of interannual modes of variability (i.e., IOD and ENSO) was removed, the forecast still performed well, implying that forecast quality does not solely depend on these modes of variability.
The CRPSS values also vary, being worse in the positive phase and better in the negative phase. Just like for the MJO and Kelvin wave, the wet phase increases the frequency and intensity of rainfall, which in turn increases the error between the reforecasts and observations on days with enhanced rainfall. These results despite the small sample, suggest that the IOD leads to variation of the reforecasts skill.

6.4 Discussion and Conclusions

This chapter analysed the role of climate drivers, specifically, EWs, the MJO and IOD on the skill of rainfall forecasts over EEA. The ability of the drivers to modulate rainfall was first analysed. Then the impact of the drivers on reforecasts skill was considered. IMERG rainfall accumulations and skill computed in the previous chapter were stratified based on the different phases of the climate drivers. MJO and EWs were identified by filtering IMERG daily rainfall using the wavenumber-frequency spectrum method of Wheeler and Kiladis (1999). Then the local phases at two locations in EEA were defined.
The influence of IOD on the rainfall and skill of reforecast was also tested. The results are limited to checking the variation of the rainfall and skill already computed in the previous chapter and not doing a detailed diagnostic analysis of regime-dependent skill which is beyond the scope of this project. Findings from this analysis are summarised as follows:

1. EWs activity prevails over EEA at different time scales and with varying effects. The waves modulate rainfall intensity and frequency at different levels, with the most prominent being MJO and Kelvin waves. The fact that these two are most dominant over EEA is in line with theory. The MRG and ER are antisymmetric and relatively further away from the equator (e.g., Matsuno, 1966; Wheeler and Kiladis, 1999) and therefore, compared to the MJO and Kelvin wave, have relatively smaller impacts in most areas in EEA. The MJO modulate the rainfall up to ~5 mmd^{-1} mostly over the Eastern sector of the study domain, i.e., UG and EAC regions. This is also in line with previous studies like Pohl and Camberlin (2006a),

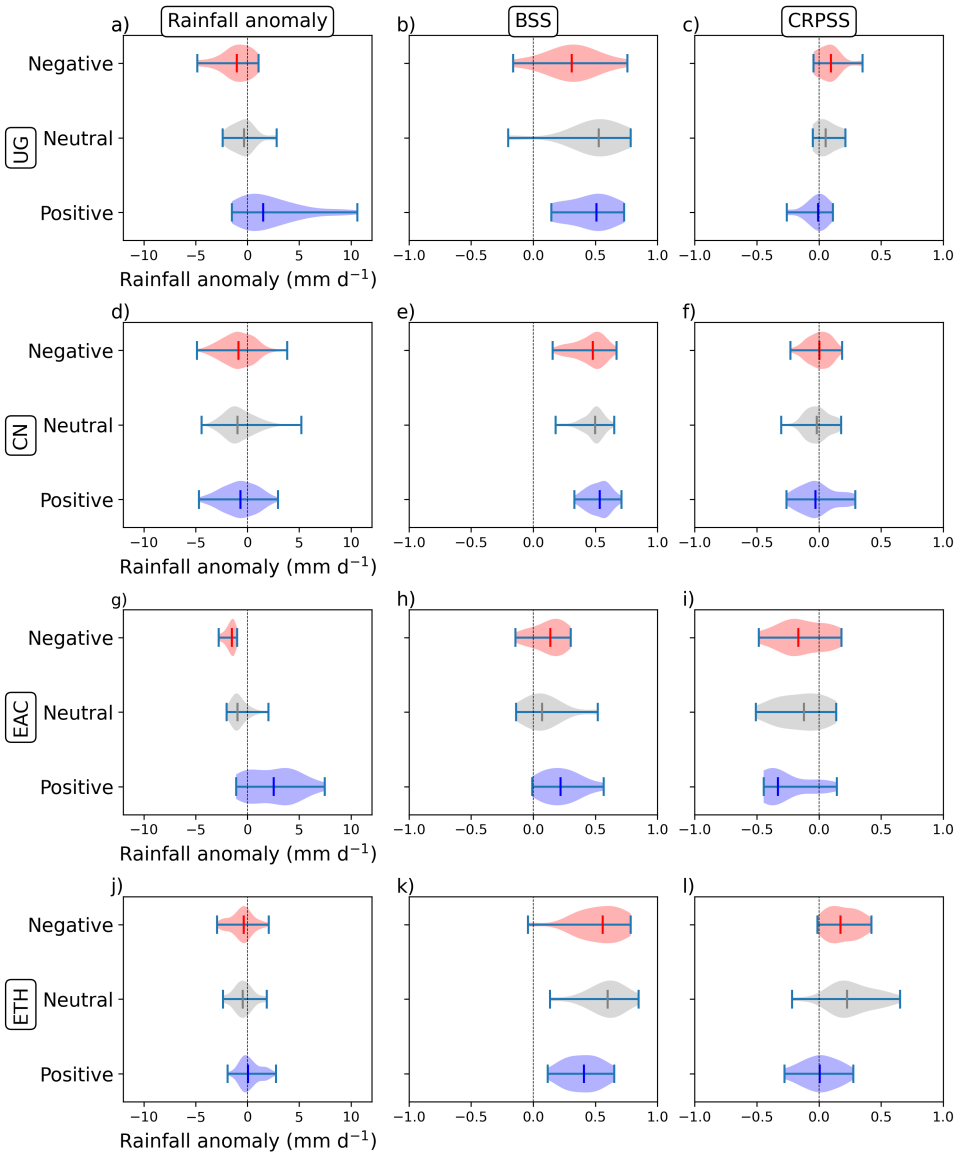


Figure 6.8: Variation of rainfall anomalies (first column) and raw reforecasts skill (BSS in the second column and CRPSS in the last column) by IOD. The modulation was analysed for the four regions in EEA.

Pohl and Camberlin (2006b) and Berhane and Zaitchik (2014) among others which also showed that MJO mainly enhances convection over the East African highlands. The Kelvin wave on the other hand is more dominant over the Congo basin region. The wave has a stronger impact on rainfall, causing variations of ~7 mmd^{-1}. Additionally, the wave brings similar effects on the rainfall in all seasons, unlike MJO which is mostly inactive in JJA season. This rainfall modulation by the Kelvin wave has also been seen in previous studies (e.g., Mekonnen and Thorncroft, 2016; Ayesiga et al., 2021, 2022). The fact that the Kelvin wave is active in all seasons was also shown in Ayesiga et al. (2022).

2. The MJO and Kelvin waves influence the skill of reforecasts over EEA. Despite the limited sample size, it can be seen that the skill scores for rainfall detection (i.e., BSS) are higher during the wet phases of both the MJO and the Kelvin wave. This could be linked to the fact that during the wet phase, the frequency of rainfall occurrence is overall increased for a given sample size of dates. Hence the hit rate is increased which translates to higher BSS values, with the difference between the wet and dry phase being ~0.2. Similar results were seen in Specq and Batté (2022). Other studies have also cited MJO as a source of predictability in EEA (Macleod et al., 2021; de Andrade et al., 2021). It has also been recommended that continuously monitoring the Kelvin wave activities over the region may be beneficial to forecast accuracy (Ayesiga et al., 2022).
   During the dry phases (1–3 and 8) of the MJO, rainfall is reduced in most parts of EEA. The dry conditions mean the bias is small and hence the CRPSS values are higher. On the contrary, because of lower hit rate, the BSS values are lower.

3. The IOD also modulates rainfall over EEA. The modulation is lower than that seen for MJO and Kelvin wave, hardly reaching 4 mmd^{-1}. The fact that only two years for each of the three phases of the IOD were available, and the IOD indices in the years considered here were not the strongest, may have obscured the influence of the IOD. However, the fact that even with these limitations, for Uganda and East African Coast, there was some level of modulation, which supports the general agreement that IOD modulate rainfall in the region (Black et al., 2003; Hirons and Turner, 2018; Wainwright et al., 2021; de Andrade et al., 2021).
   In regards to the reforecasts, although not obvious, over Uganda and the East African Coast, there is a tendency of better BSS values during the positive phase. The CRPSS values on the other hand are worse during this phase. These findings confirm what is observed with the MJO and Kelvin waves, that the wet phases of the climate drivers improve rainfall detection while increasing the bias. Because the IOD develops slowly over time, it is usually well predicted by weather forecasting systems, and this could be another possible explanation for improved skill.

This chapter has highlighted the influence of climate drivers on rainfall variability over EEA which in turn drives variability in the skill of the model forecasts. It seems that when the positive phase of the driver prevails, the models have a better rainfall prediction ability but the bias for rainfall amount increases. This information can be useful to forecasters. Moreover, some of the drivers analysed here, MJO and Kelvin waves propagate eastwards. Therefore, monitoring their progression should be done to increase the forecast skill of the region. Additionally, the results suggest that the importance of the different drivers varies spatio-temporally with the MJO dominant over Uganda and the East African highlands, the Kelvin wave dominant over Congo and

the IOD influencing rainfall and forecast skill over the coastal areas and Uganda region. This study only highlighted modulation of rainfall and variation of skill when different phases of three weather phenomenon prevail. The study does not analyse in details the mechanisms responsible for the improved or reduced skill of the reforecasts in these phases. Future work could be to do a detailed diagnostic analysis of rainfall and skill variation as a result of these climate drivers. Moreover, since these drivers do not operate in singularity (e.g., de Andrade et al., 2021), doing composites of more than one climate driver as done in Vigaud et al. (2018) and also analysing other drivers like the Congo air mass may yield more robust results. Finally, the nature of reforecsats, issued only twice in a week means that only a one third of total number of days in a year are available for analysis. If these are further selected by season and state of the climate driver (selected only if active), the sample size shrinks even further.

7. Conclusions and Outlook

The EEA region is a region which is heavily reliant on rain-fed agriculture for the socio-economic livelihood of its population. The region is highly vulnerable to weather and climate extremes which, in recent decades, have increased both in intensity and frequency. This makes weather and climate information very crucial for this part of the world. However, it is also true that the skill of the forecasts is poor and their uptake is low. Although there are a lot of efforts toward reversing this scenario, for now, the enormous potential of weather and climate information has not been realised in Africa. This work therefore aimed at contributing to efforts to improve forecasts by validating the forecasts over the region using the best available observation data sets, IMERG and rain gauges. In particular, a comprehensive analysis of the skill of SREs and reforecasts at different spatio-temporal aggregation was performed. Here, a summary of the main findings and suggestions for future work are given .

7.1 Conclusions

The first aim of this work was to find the best SRE for EEA. Three questions were asked to address this aim. These questions are answered in chapter four and are summarised as follows:

RQ 1a **What is the best performing SRE over EEA at daily, pentadal, decadal, monthly and annual time scales?**

First a thorough quality check of gauge data from around Uganda. This quality control resulted into a unique clean and consistent rainfall data set composed of 36 stations for the period 2002–2018, which enabled a comprehensive evaluation of the SREs.

On the annual and seasonal time scales, CHIRPS and MSWEPv2.2 were the best products. In general, all the SREs reproduce the annual patterns and seasonal cycle of rainfall, albeit overestimating the annual totals in the range of 1–9%. This result supports findings in previous studies such as Camberlin et al. (019a), Diem et al. (2014) and Asadullah et al. (2008). The reason for this good agreement between SREs and ground observations

can be explained by the fact that the seasonal migration of the ITCZ is well-captured in the products (e.g., Nicholson, 2017; Seregina et al., 2018). Additionally, all the products use gauge observation for calibration (Awange et al., 2015; Dinku et al., 2018).

At the daily, pentadal and decadal time scales, IMERG showed the best performance followed by MSWEPv2.2, TMPA and then CHIRPS. Dezfuli et al. (017a) also showed that IMERG was better than its predecessor TMPA. They explained that the good performance of IMERG is due to a better spatio-temporal resolution. It should be noted that, despite its generally better performance, IMERG overestimates rainfall over regions where convective rainfall dominates, e.g., Lake Victoria. This is because, the PMW estimates which are the major component used in IMERG, tend to overestimate convective rainfall (Tian et al., 2009).

RQ 1b Does the performance of the SREs vary spatially over EEA?

The performance of the SREs substantially varies with location, mainly due to topography. For example, over southwestern Uganda, a hilly region punctuated by valleys, the SREs struggled. Over the flatter areas of northern Uganda, they performed better. The fact that SREs struggle over high altitude areas like mountains has also been reported in Diem et al. (2014) and Monsieurs et al. (2018). In southwestern Uganda particularly, a higher false alarm rate accounted for the poor skill due to the region being located in a rain shadow. Secondly, and perhaps a salient reason to many people, the gauge density used for calibration, the weight assigned to these gauges and their location, influence the spatial skill distribution, as seen mostly in the results of MSWEPv2.2. In regions with higher gauge densities, and where a higher gauge weight relative to the other data sources was used when creating MSWEPv2.2, it was significantly better than where few stations were available or low weights were applied to the gauge data. Therefore, it is very important for SREs creators to supplement the rainfall data with supplementary information regarding the data inputs used to create the SRE.

RQ 1c Are the SREs capable of retrieving extreme rainfall events over EEA and, if so, which is the best product?

The SREs struggle to retrieve extreme rainfall over the region. Indeed, all the products, with the exception of the early version of IMERG, miss 80% of extreme daily rainfall accumulation. This substantially affects their suitability for application like modelling return periods of very extreme events. Products like MSWEPv2.8, CHIRPS and the early version of

TMPA are not recommended for this purpose in this region given that they consistently ranked worst for ability to retrieve extreme rainfall events. Thiemig et al. (2012) and Monsieurs et al. (2018) found similar results over African river basins and western Uganda, respectively. Possible explanations for the poor performance of the SREs in retrieving extreme rainfall are; a) the fact that passive microwave sensors do not record continuously over a particular region, hence they may miss short-lived intense rainfall due to the infrequent overpass, and b) SREs are often smoothed fields and therefore the extreme rainfall may be averaged out. The fact that the early version of IMERG performed best in retrieving extreme rainfall makes a case for studies like Bitew and Gebremichael (2011), which suggested that calibration, often done for the final version of SREs, does introduce errors. Most times the calibration is done at monthly time scales before rescaling to daily values, which could be a source of errors.

From the above answers, it is clear that validation of SREs before they are used is crucial. As demonstrated above, no single product consistently outperformed the others in all regions of the study domain and/or time scales. At the shorter aggregation time scales, IMERG generally outperformed all the other SREs assessed here, hence it is chosen for validating the ECMWF reforecasts. The fact that it also ranked best for extreme rainfall events makes it an even stronger candidate.

In chapter five, results answering questions aimed at achieving the second aim of this work were presented. The findings are:

RQ 2a Does the ECMWF deterministic reforecast correctly estimate the intensity of rainfall over EEA?

The deterministic reforecast, here, taken to be the ensemble median, overestimated the next day rainfall accumulation over raised areas, especially in the rainy seasons. The overestimation increases with a higher rainy day threshold, meaning that the error grows as the rainfall intensities become more extreme. Over low-lying regions such as the Congo basin, the reforecasts underestimated rainfall. Similar findings but for weekly time scales, were reported in de Andrade et al. (2021) over the same region and by Stellingwerf et al. (2021) over the Ethiopian highlands. However, over the raised areas of EEA, IMERG also struggled retrieving rainfall. This may imply that the overestimation by the reforecast may partly be attributable to errors in the observation data set. The fact that the overestimation happens during the wet season is considered positive, as it demonstrates that the model puts rainfall in the right places, albeit with some errors.

RQ 2b Do the raw ensemble ECMWF reforecasts predict the correct occurrence of a rainy day and rainfall amounts and are they better than a reference forecast?

Over most land grid-points, the reforecasts correctly predicted the next day rainfall based on a 0.2 mm rainfall occurrence threshold. There is a 30% improvement in the detection rate compared to a reference forecast, EPC. The fact that the ECMWF model has positive skill in this region has been shown by other studies (e.g., de Andrade et al., 2021; Endris et al., 2021; Stellingwerf et al., 2021). However, when tested using the DM test, the performance of the model was not significantly different from that of EPC. Just like the deterministic reforecast, the performance varies with season, MAM and DJF posting the best BSS and CRPSS values, followed by SON, and the worst performance seen in JJA. The model seems to perform better in predicting rainfall occurrence compared to rainfall amounts, i.e., BSS > CRPSS values for the same grids. The performance degrades with lead time as expected. This is because, with longer lead times, the model loses the memory gained from the initial conditions. Moreover, the model also performs better for longer rainfall accumulations (i.e., 48-hour and 120-hour accumulations) since any error due to temporal mismatches are reduced. This level of skill in the models can be attributed to the better representation of climate drivers like MJO (de Andrade et al., 2021; Vitart et al., 2017). However for some grids, the performance of the reforecast was worse than that of the reference forecast. Such performance supports findings of Haiden et al. (2012) and Vogel et al. (2018) which showed poor skill in global models. The poor model performance has mainly been linked to inadequacy of the convection parameterization (e.g., Marsham et al., 2013; Birch et al., 2014; Vogel et al., 2018) and the stochasticity of convection in regions like the Congo basin (Satheesh et al., 2023).

RQ 2c Do the reforecasts have any systematic biases and can postprocessing using IDR alleviate the inadequacy?

The reforecasts have biases and are overconfident. Postprocessing using IDR substantially reduced the miscalibration. For many grid-points, the improvement was 50% over the reference forecast, especially in the two rainy seasons and DJF dry season in both the BS and CRPS. Counting the number of grids with positive skill reveals that the postprocessing improved the BSS (CRPSS) by 82% (48%). Similar improvements were seen by Vogel et al. (2020) over West Africa. Postprocessing using the rain gauge data shows similar results.

RQ 2d **Are there difference in the performance of the reforecasts over the different topographical features of EEA?**
The skill was generally better over high altitudes compared to the low-lying regions. There was a northeast-southwest orientation of enhanced skill which coincides with the location of the Ethiopian Highlands, Uganda and the southern Congo/western Tanzania. These regions have a relatively higher altitude in the study domain. This performance suggests that the model correctly represents orographically triggered rainfall. Studies like Pohl and Camberlin (2006a) also suggest that precipitation over these regions is linked to climate drivers like the MJO, which increase the predictability of the rain in these regions. Over the low-lying regions, specifically the Congo Basin, the model has a very low discrimination ability. Moreover, rainfall here is very convective and stochastic (Satheesh et al., 2023), hence the model struggles to correctly represent it. This, together with the well-known convection parameterization struggles of numerical weather prediction models, magnifies the problem. Again, the analysis against gauges over three location in Uganda reinforced this result.

Clearly there is positive skill for many land grid points, but also regions and times with very poor skill. This underlines the variable nature of skill in the tropics. The next logical question is: what leads to skillfulness or no-skill in the reforecast? As discussed in the literature (chapter two), there are several climate drivers which modulate rainfall in EEA. These, despite bringing about variability, also offer opportunity for forecasting. Therefore, the last aim of this work was to assess the influence of different weather regimes on the skill of forecasts over EEA. Specifically, the role of EWs including the MJO. The research questions to achieve this third goal were answered in chapter six. A summary is given below:

RQ 3a **Do EWs and the MJO modulate rainfall over EEA and by how much?** The rainfall over EEA is variable at multiple time scales. The EWs, specifically the Kelvin wave and MJO, play a key role in bringing about this variability. The MJO is mainly strong in the eastern part of the study domain, i.e., Uganda and the East African coast where, it modulates rainfall by ~5 mmd^{-1} in either direction. The Kelvin wave on the other hand is more dominant over the western part of the study domain, mostly over the Congo basin, with a stronger modulation of ~7 mmd^{-1} in either direction. The Kelvin wave over Congo is active almost all year round, unlike the MJO that seems to be muted in the JJA season. The analysis of the waves over the region mostly agreed with theory laid out in Matsuno

(1966) and other related literature. The results also largely agree with previous studies that analysed the waves over this region (e.g., Pohl and Camberlin, 2006a,b; Berhane and Zaitchik, 2014; Vellinga and Milton, 2018; Vashisht and Zaitchik, 2022; Ayesiga et al., 2021; Finney et al., 2019; Specq and Batté, 2022).

RQ 3b Do EWs and the MJO influence the skill of the ECMWF reforecasts over EEA? The activity of the climate drivers influences the skill of the reforecasts. During the wet phases of the MJO and Kelvin waves, both the BSS and CRPSS are better. This is because, with enhanced rainfall frequency, the hit rate is improved which translates to higher BSS during the wet phases (Specq and Batté, 2022). During the dry phases, the picture is reversed with the scores being worse. However, the influence of these waves is variable over EEA, with MJO (Kelvin waves) being dominant over Uganda and EAC (Congo). This in agreement with past studies which showed MJO is most dominant over the East African highlands (e.g., Pohl and Camberlin, 2006a). Ayesiga et al. (2022) previously showed the strong influence of Kelvin waves over the Congo.

RQ 3c Does the IOD modulate rainfall over EEA? What is its influence on the skill of ECMWF reforecasts? The IOD modulate rainfall, especially over Uganda and the East African Coast areas. During the positive phase, the IOD brings more rains to these regions and consequently better BSS and worse CRPSS values. This is consistent with previous studies (e.g., Black et al., 2003; Wainwright et al., 2021). In the negative phase, the reverse is true. These regions are drier and the BSS tends to get worse while the CRPSS improves. However, the modulation for IOD is less pronounced than that seen for the MJO and Kelvin wave.

7.2 Outlook

This thesis provided a comprehensive verification of ECMWF reforecasts over EEA, verifies SREs, and quality checked rain gauge data for selected stations over Uganda. Additionally, the study explores how the performance of reforecats varies when climate drivers prevail over the domain. The results shed light on the current state of forecast quality over EEA, giving a basis for improvement of skill. Although the verification was comprehensive, some aspects of the forecasts skill could not be analysed as they were beyond the scope of this study. Therefore, the following are suggested for future work aimed at improving forecast quality:

Analysis based on finer resolution data should be done to assess variability at more localised scales. This work is based on reforecsast data that are initialised daily and

archived on 1.5°×1.5°grid-spacing and initialised only twice weekly. Moreover all data sets in the S2S database have fewer ensemble members (11 for ECMWF) compared to the operational forecasts. These few members limit the assessment of uncertainty in the reforecasts. Therefore, the spatio-temporal resolution does not allow identification of small-scale variation. With the computing resources getting cheaper, it is likely that the resolution of reforecast data in the S2S data base could get higher. Another option is to repeat the study using operational forecasts which have a higher spatio-temporal resolution (0.25°×0.25°) and more members (currently 100) to better assess uncertainty. The coarse resolution limited analysis with gauge data. Only four grid cells had a reasonable number (in this case, a minimum of four) of gauges within them, validation of the reforecasts with gauges was performed at only four grid points. An option could be a field campaign where several gauges could be set up in order to increase the gauge density for more robust results. Given the high variability of the rainfall and reforecast skill, an analysis with higher resolution data would allow even finer forecast differences to be identified, enabling improved forecasts for local weather.

The idea of multi-model forecast ensembles is one which has been suggested to improved forecast skill (e.g., Stellingwerf et al., 2021). In this study, only ECMWF reforecasts have been analysed. Although the reasoning behind this choice is sound, i.e., it has been found to outperform forecast products from other forecasting centres in the region (e.g., de Andrade et al., 2021; Stellingwerf et al., 2021; Endris et al., 2021), it may not be the best for all the desirable forecast attributes. Moreover, the small ensemble size problem is dealt with if a multi-model ensemble is used. This would increase the ensemble spread and reduce uncertainties. The obvious limitation to this approach would be how to reconcile the different model characteristic, such as, different ensemble sizes, different forecasting techniques used (e.g., some forecasts are produced on the fly while others are fixed) and temporal resolution differences. However, reasonable solutions can be found. For example in de Andrade et al. (2021), the smaller ensemble size (7) of the NCEP model compared to ECMWF (11) was solved by including perturbed ensemble members from one day after forecast initialization.

Further work can also consider using more than the one postprocessing method used in this study. In this study, the advantage of postprocessing was shown at all time scales and regions. A recent simple technique, IDR, was used mostly due to its simplicity and it did not require prior determining of predictors (Henzi et al., 2021). Moreover, it was shown to be competitive when compared to other postprocessing methods (Maier-Gerber et al., 2021; Schulz and Lerch, 2022). However, it is not perfect, for example it would not do well if the data sample is small. IDR needs a long training data set to train the model. Therefore, testing other postprocessing approaches, for example, EMOS or Member-by-Member postprocessing is recommended as future work. Additionally, machine learning approaches can be tested over this region. In recent times, machine

learning in weather forecasting has rapidly developed, supported by cheaper computing resources. Schulz and Lerch (2022) showed that machine learning approaches resulted in the best scores compared to classical postprocessing methods over Europe.

In this study, the regime-dependant skill analysis only focused in assessing if the skill varies depending on specific prevailing climate drivers, specifically, MJO, Kelvin waves and the IOD. The first step for further research would be to expand the climate drivers by considering other climate drivers in the region. For example, since IOD and ENSO covary, ENSO could be used to extend the assessment of the these modes of variability to DJF when the impact of IOD has vanished. The Congo airmass is another candidate at short to medium-range time scales, as well as the influence of topographical features (Lake Victoria and the highlands). The current study already highlighted the role of topography on skill, with better skill generally over higher altitude areas compared to the low-lying areas. However, the differences were not quantified. A detailed study of these features could be done. Lake Victoria has been shown to be a significant feature in terms of weather modification mainly due to its strong lake-land breezes (Woodhams et al., 2018). Exactly how it influences the skill could be analysed. Even for the drivers analysed in this thesis, a diagnostic analysis was not done. Future work should focus on looking at what mechanism and dynamics of these drivers lead to the reduced or improved skill during specific phases. Finally, although the influence of climate drivers was assessed individually, they do not occur exclusively. For example, most times and ENSO and IOD occur at the same period to cause devastating effects as was the case in 1998. In this study, all the EWs and MJO were present in the MAM season of 2012, which could allow generating composites of combinations of these drivers.

As a final step, most of the work shown in this thesis and other research carried out in universities and research centres is not effective if not shared with the forecast users, especially the National Meteorological and Hydrological Services (NMHS) in Africa. Clearly there is predictability of weather over Africa, but this is localised. Forecasters can benefit from this information and improve their forecasts. They can also learn the systematic biases in the model products they use. Moreover, this study also highlighted the data challenge over Africa which may encourage the NMHSs to be more willing to share their data, which more often than not will not reach the GTS in time to be assimilated in the model.

Bibliography

Ageet, S., Dink, A. H., Maranan, M., Diem, J. E., Hartter, J., Ssali, A. L., and Ayabagabo, P. (2022). Validation of satellite rainfall estimates over Equatorial East Africa. *J. Hydrometeor.*, 23:129–151.

Ageet, S., Fink, H. A., Maranan, M., and Schulz, B. (2023). Predictability of rainfall over Equatorial East Africa in the ECMWF ensemble hindcasts on short to medium-range time scales. *Weather and Forecating*.

Asadullah, A., McIntyre, N., and Kigobe, M. (2008). Evaluation of five satellite products for estimation of rainfall over Uganda. *Hydrol. Sci. J.*, 53:1137–1150.

Awange, J., Ferreira, V., Forootan, E., Khandu, K., Andam-Akorful, S., Agutu, N., and He, X. (2015). Uncertainties in remotely sensed precipitation data over Africa. *International Journal of Climatology*, 36:303–323.

Ayesiga, G., Holloway, C. E., Williams, C. J., Yang, G.-Y., and Ferrett, S. (2021). The observed synoptic scale precipitation relationship between Western Equatorial Africa and Eastern Equatorial Africa. *International Journal of Climatology*, 41(S1):E582–E601.

Ayesiga, G., Holloway, C. E., Williams, C. J. R., Yang, G.-Y., Stratton, R., and Roberts, M. (2022). Linking Equatorial African Precipitation to Kelvin Wave Processes in the CP4-Africa Convection-Permitting Regional Climate Simulation. *Journal of the Atmospheric Sciences*, 79(5):1271 – 1289.

Bahaga, T. K., Mengistu Tsidu, G., Kucharski, F., and Diro, G. (2014). Potential Predictability of the SST-Forced Equatorial East African Short Rain Interannual Variability in the 20th Century. *Quarterly Journal of the Royal Meteorological Society*, 141:16–26.

Barnston, A., Tippett, M., L'Heureux, M., Li, S., and DeWitt, D. (2012). Skill of Real-Time Seasonal ENSO Model Predictions during 2002-11: Is Our Capability Increasing? *Bulletin of the American Meteorological Society*, 93:631–651.

Basalirwa, C. P. K. (1995). Delineation of Uganda into climatological rainfall zones using the method of principal component analysis. *International Journal of Climatology*, 15(10):1161–1177.

Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., Van Dijk, A. I., McVicar, T. R., and Adler, R. F. (2019). MSWep v2 Global 3-hourly 0.1° precipitation: Methodology and quantitative assessment. *Bulletin of the American Meteorological Society*, 100(3):473–500.

Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., Van Dijk, A. I., McVicar, T. R., and Adler, R. F. (2021a). MSWEPv2.8 3-hourly 0.1° × 0.1° historical data. GloH2O. Accessed 1 March 2021, `http://www.gloh2o.org/mswep/`.

Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., Van Dijk, A. I., McVicar, T. R., and Adler, R. F. (2021b). MSWEPv2.8 technical documentation. GloH2O. Accessed 4 May 2021, `http://www.gloh2o.org/mswep/`.

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., and Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, 5.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. Roy. Stat. Soc.*, 57B:289–300.

Berhane, F. and Zaitchik, B. (2014). Modulation of Daily Precipitation over East Africa by the Madden–Julian Oscillation*. *Journal of Climate*, 27.

Birch, G. E., Parker, D. J., Marsham, J. H., Copsey, D., and Garcia-Carreras1, L. (2014). A seamless assessment of the role of convection in the water cycle of the West African Monsoon. *J. Geophys. Res. Atmos.*, 119:2890–2912.

Bitew, M. and Gebremichael, M. (2011). Assessment of high-resolution satellite rainfall for streamflow simulation in medium watersheds of the East African highlands. *Hydrology and Earth System Sciences - HYDROL EARTH SYST SCI*, 15:1147–1155.

Bjerknes, J. (1966). A possible response of the atmospheric Hadley circulation to equatorial anomalies of ocean temperature. *Tellus*, 18(4):820–829.

Bjerknes, J. (1969). Atmospheric Teleconnections from the Equatorial Pacific. *Monthly Weather Review*, 97(3):163 – 172.

Black, E., Slingo, J., and Sperber, K. R. (2003). An Observational Study of the Relationship between Excessively Strong Short Rains in Coastal East Africa and Indian Ocean SST. *Monthly Weather Review*, 131(1):74 – 94.

Bommier, E. (2014). Peaks-Over-Threshold Modelling of Environmental Data. Technical report.

Bormann, N., Lawrence, H., and Farnan, J. (2019). Global observing system experiments in the ECMWF assimilation system.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, 78:1–3.

Cafaro, C., Woodhams, J. B., Stein, H. M. T., Birch, E. C., Webster, S., Bain, L. B., Hartley, A., Clerke, S., Ferrett, S., and Hill, P. (2021). Do convective-permitting ensembles lead to more skillful short-range probabilistic rainfall forecasts over tropical East Africa? *Wea. Forecasting*, 36:697–716.

Camberlin, P., Barraud, G., Bigot, S., Dewitte, O., Imwangana, F., Maki Mateso, J.-C., Martiny, N., Monsieurs, E., Moron, V., Pellarin, T., Philippon, N., Sahani, M., and Samba, G. (2019a). Evaluation of remotely sensed rainfall products over Central Africa. *Quarterly Journal of the Royal Meteorological Society*, 145:2115–2138.

Camberlin, P., Gitau, W., Kiladis, G., Bosire, E., and Pohl, B. (2019b). Intraseasonal to Interannual Modulation of Diurnal Precipitation Distribution Over Eastern Africa. *Journal of Geophysical Research: Atmospheres*, 124(22):11863–11886.

Camberlin, P., Gitau, W., Planchon, O., Dubreuil, V., Funatsu, B. M., and Philippon, N. (2018). Major role of water bodies on diurnal precipitation regimes in Eastern Africa. *International Journal of Climatology*, 38(2):613–629.

Cattani, E., Merino, A., and Levizzani, V. (2016). Evaluation of Monthly Satellite-Derived Precipitation Products over East Africa. *Journal of Hydrometeorology*, 17(10):2555 – 2573.

CHC (2023). Updated climate forecasts continue to suggest below-normal 2023 March-April-May rains likely in Eastern Horn of Africa. Accessed 14 June 2023, `https://blog.chc.ucsb.edu/?p=1263`.

Chen, W. Y. (1982). Assessment of Southern Oscillation Sea-Level Pressure Indices. *Monthly Weather Review*, 110(7):800 – 807.

Coz, C. L. and van de Giesen, N. (2020). Comparison of Rainfall Products over Sub-Saharan Africa. *Journal of Hydrometeorology*, 21(4):553 – 596.

de Andrade, F. M., Young, M. P., Macleod, D., Hirons, L. C., Woolnough, S. J., and Black, E. (2021). Subseasonal precipitation prediction for Africa: Forecast evaluation and sources of predictability. *Wea. Forecasting*, 36:265–284.

Dezfuli, A., Ichoku, C., Huffman, G., Mohr, K., Selker, J., van de Giesen, N., Hochreutener, R., and Annor, F. (2017b). Validation of IMERG precipitation in Africa. *Journal of Hydrometeorology*, 18:2817–2825.

Dezfuli, A., Ichoku, C., Mohr, K., and Huffman, G. (2017a). Precipitation Characteristics in West and East Africa from Satellite and in Situ Observations. *Journal of Hydrometeorology*, 18:1799–1805.

Diebold, X. F. and Mariano, S. R. (1995). Comparing predictive accuracy. *J. Bus. Econ. Stat.*, 13:253–263.

Diem, J., Konecky, B., Salerno, J., and Hartter, J. (2019). Is equatorial Africa getting wetter or drier? Insights from an evaluation of long-term, satellite-based rainfall estimates for western Uganda. *International Journal of Climatology*, 39:3334–3347.

Diem, J. E., Hartter, J., Ryan, S. J., and Palace, M. W. (2014). Validation of satellite rainfall products for western Uganda. *J. Hydrometeor.*, 15:2030–2038.

Dimitriadis, T., Gneiting, T., and Jordan, A. I. (2021). Stable reliability diagrams for probabilistic classifiers. *Proc. Natl. Acad. Sci.(USA)*, 118(8 e2016191118).

Dinku, T. (2019). Chapter 7 - Challenges with availability and quality of climate data in Africa. In Melesse, A. M., Abtew, W., and Senay, G., editors, *Extreme Hydrology and Climate Variability*, page 71–80. Elsevier.

Dinku, T., Ceccato, P., Cressman, K., and Connor, S. J. (2010a). Evaluating Detection Skills of Satellite Rainfall Estimates over Desert Locust Recession Regions. *Journal of Applied Meteorology and Climatology*, 49(6):1322 – 1332.

Dinku, T., Connor, S., and Ceccato, P. (2010b). *Comparison of CMORPH and TRMM-3B42 over mountainous regions of Africa and South America*, pages 193–204.

Dinku, T., Funk, C., Peterson, P., Maidment, R., Tadesse, T., Gadain, H., and Ceccato, P. (2018). Validation of the CHIRPS Satellite Rainfall Estimates over Eastern of Africa: Validation of the CHIRPS Satellite Rainfall Estimates. *Quarterly Journal of the Royal Meteorological Society*, 144:292–312.

Dunning, C. M., Black, E., and Allan, R. P. (2018). Later Wet Seasons with More Intense Rainfall over Africa under Future Climate Change. *Journal of Climate*, 31(23):9719–9738.

Dunning, C. M., Black, E. C. L., and Allan, R. P. (2016). The onset and cessation of seasonal rainfall over Africa. *Journal of Geophysical Research: Atmospheres*, 121(19):11,405–11,424.

Ebert, E. (2007). *Methods for Verifying Satellite Precipitation Estimates*, volume 28, pages 345–356.

Eggermont, H., Van Damme, K., and Russell, J. (2009). *Rwenzori Mountains (Mountains of the Moon): Headwaters of the White Nile*, pages 243–261.

Endris, H. S., Hirons, L., Segele, Z. T., Gudoshava, M., Woolnough, S., and Artan, G. A. (2021). Evaluation of the skill of monthly precipitation forecasts over the Greater Horn of Africa. *Wea. Forecasting*, 36:1275–1298.

Engel, T., Fink, A., Knippertz, P., Pante (né Gläser), G., and Bliefernicht, J. (2017). Extreme Precipitation in the West African Cities of Dakar and Ouagadougou: Atmospheric Dynamics and Implications for Flood Risk Assessments. *Journal of Hydrometeorology*, 18:2937–2957.

Faccani, C., Rabier, F., Fourrié, N., Agusti-Panareda, A., Karbou, F., Moll, P., Lafore, J.-P., Nuret, M., Hdidou, F., and k, O. B. (2009). The Impacts of AMMA Radiosonde Data on the French Global Assimilation and Forecast System. *Weather and Forecasting*, 24(5):1268 – 1286.

Ferrett, S., Yang, G.-Y., Woolnough, S. J., Methven, J., Hodges, K., and Holloway, C. E. (2020). Linking extreme precipitation in Southeast Asia to equatorial waves. *Quarterly Journal of the Royal Meteorological Society*, 146(727):665–684.

Findlater, J. (1969). A major low-level air current near the Indian Ocean during the northern summer. *Quarterly Journal of the Royal Meteorological Society*, 95(404):362–380.

Finney, D., Marsham, J., Walker, D., Birch, C., Woodhams, B., Jackson, L., and Hardy, S. (2019). The effect of westerlies on East African rainfall and the associated role of tropical cyclones and the Madden-Julian Oscillation. *Quarterly Journal of the Royal Meteorological Society*, 146:647–664.

Flohn, H. and Burkhardt, T. (1985). Nile runoff at Aswan and Lake Victoria; an example of a discontinuous climate time series. *Z. Gletschkd. Glazialgeol.*, 21:125–130.

Flohn, H. and Fraedrich, K. (1966). Tagesperiodische Zirkulation und Niederschlagsverteilung am Victoria-See (Ostafrika). *Meteor. Rundsch.*, 19:157–165.

Funk, C., Harrison, L., Shukla, S., Pomposi, C., Galu, G., Korecha, D., Husak, G., Magadzire, T., Davenport, F., Hillbru ner, C., Eilerts, G., Zaitchik, B., and Verdin, J. (2018). Examining the role of unusually warm Indo-Pacific sea-surface temperatures in recent African droughts. *Quarterly Journal of the Royal Meteorological Society*, 144(S1):360–383.

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*, 2:150066.

Gill, A. E. (1980). Some simple solutions for heat-induced tropical circulation. *Quarterly Journal of the Royal Meteorological Society*, 106(449):447–462.

Gneiting, T. and Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application*, 1(1):125–151.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *J. Amer. Stat. Assoc.*, 102:359–378.

Gneiting, T., Raftery, A. E., Westveld, A. H., and Goldman, T. (2005). Calibrated probablistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Wea. Rev.*, 133:1098–1118.

Griffiths, J. F. and Landsberg, H. E. (1972). *World survey of climatology. 10. Climates of Africa.* Elsevier.

Haiden, T., Rodwell, M. J., Richardson, D. S., Okagaki, A., Robinson, T., and Hewson, T. (2012). Intercomparison of global model precipitation forecast skill in 2010/11 using SEEPS score. *Mon. Wea. Rev.*, 140:2720–2733.

Ham, Y.-G., Kim, J.-H., and Luo, J.-J. (2019). Deep learning for multi-year enso forecasts. *Nature*, 573:568–572.

Haoyu, W., Hu, S., and Li, X. (2023). An interpretable deep learning enso forecasting model.

Hartigan, J. A. and Coauthors (1999). The Global Land One-Kilometer Base Elevation (GLOBE) digital elevation model, version 1.0. Accessed 23 July 2021, `https://www.ngdc.noaa.gov/mgg/topo/globe.html`.

Hartigan, J. A. and Wong, M. A. (1979). A K-means clustering algorithm. *J. Roy. Stat. Soc.*, 28:100–108.

Henzi, A., Ziegel, J. F., and Gneiting, T. (2021). Isotonic distributional regression. *J. Roy. Stat. Soc.*, 83:963–993.

Hersbach, H., Bill, B., Berrisford1Shoji, H., Horányi, A., Joaquín, M.-S., Julien, N., Carole, P., Raluca, R., Dinand, S., Adrian, S., Cornel, S., Saleh, A., Xavier, A., Gianpaolo, B., and Coauthor (2020). The ERA5 global reanalysis. *Q. J. R. Meteorol.*, 146:1999–2049.

Hirons, L. and Turner, A. (2018). The Impact of Indian Ocean Mean-State Biases in Climate Models on the Representation of the East African Short Rains. *Journal of Climate*, 31(16):6611 – 6631.

Hogan, E., Shelly, A., and Xavier, P. (2014). The observed and modelled influence of the Madden-Julian Oscillation on East African rainfall: The influence of the MJO on East African rainfall. *Meteorological Applications*, 22:459–469.

Huang, P., Chou, C., and Huang, R. (2012). The activity of convectively coupled equatorial waves in CMIP3 global climate models. *Theoretical and Applied Climatology*, 112:697–711,.

Huffman, G., , and Bolvin, D. T. (2018). TRMM and other data precipitation set documentation. NASA TRMM Doc. Technical report. `https://docserver.gesdisc.eosdis.nasa.gov/public/project/GPM/3B42_3B43_doc_V7.pdf`.

Huffman, G., Adler, R., Bolvin, D., Gu, G., Nelkin, E., Bowman, K., Hong, Y., Stocker, E., and Wolff, D. (2007). The trmm multisatellite precipitation analysis (tmpa): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology - J HYDROMETEOROL*, 8:38–55.

Huffman, G., Bolvin, D. T., Braithwaite, D., Hsu, K., and Joyce, R. (2020). NASA Global Precipitation Measurement (GPM) integrated Multi-satellite Retrievals for GPM (IMERG) - Algorithm Theoretical Basis Document (ATBD). Technical report. Accessed 11 May 2020, `https://gpm.nasa.gov/sites/default/files/2020-05/IMERG_ATBD_V06.3.pdf`.

Huffman, G., Stocker, E. F., Bolvin, D. T., Nelkin, E. J., and Tani, J. (2020b). GPM IMERG Final Precipitation L3 Half Hourly 0.1°×0.1°V06 (GPM3IMERGHH). Goddard Earth Sciences Data and Information Services Center (GES DISC). Technical report. Accessed 1 April 2020, `https://doi.org/10.5067/GPM/IMERG/3B-HH/06`.

Huffman, G., Stocker, E. F., Bolvin, D. T., Nelkin, E. J., and Tani, J. (2020c). GPM IMERG Early Precipitation L3 Half Hourly 0.1° x 0.1° V06 (GPM3IMERGHHE). Goddard Earth Sciences Data and Information Services Center (GES DISC). Technical report. Accessed 1 June 2020, `https://doi.org/10.5067/GPM/IMERG/3B-HH-E/06`.

Ichiye, T. (1959). On long waves in a stratified, equatorial ocean caused by a travelling disturbance. *Deep Sea Research (1953)*, 6:16–37.

Indeje, M., Semazzi, F. H., and Ogallo, L. J. (2000). ENSO signals in East African rainfall seasons. *International Journal of Climatology*, 20(1):19–46.

IPCC (2021). *Annex IV: Modes of Variability [Cassou, C., A. Cherchi, Y. Kosaka (eds.)], booktitle = Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, editor = Masson-Delmotte, V. and Zhai, P. and Pirani, A. and Connors, S.L. and Péan, C. and Berger, S. and Caud, N. and Chen, Y. and Goldfarb, L. and Gomis, M.I. and Huang, M. and Leitzell, K. and Lonnoy, E. and Matthews, J.B.R. and Maycock, T.K. and Waterfield, T. and Yelekçi, O. and Yu, R. and Zhou, B.*, page 2153–2192. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

IPCC (2022). Summary for Policymakers [H.-O. Pörtner, D.C. Roberts, E.S. Poloczanska, K. Mintenbeck, M. Tignor, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem (eds.)]. In: Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)]. pages 3–33.

Jiang, Y., Zhou, L., Roundy, P. E., Hua, W., and Raghavendra, A. (2021). Increasing Influence of Indian Ocean Dipole on Precipitation Over Central Equatorial Africa. *Geophysical Research Letters*, 48(8):e2020GL092370.

Jones, P. W. (1999). First- and Second-order conservative remapping schemes for grids in spherical coordinates. *Mon. Wea. Rev.*, 127:2204–2210.

Judt, F. (2020). Atmospheric Predictability of the Tropics, Middle Latitudes, and Polar Regions Explored through Global Storm-Resolving Simulations. *Journal of the Atmospheric Sciences*, 77(1):257 – 276.

Kidd, C. and Huffman, G. (2011). Global precipitation measurement. *Meteorological Applications*, 18:334 – 353.

Kiladis, G. N., Wheeler, M. C., Haertel, P. T., Straub, K. H., and Roundy, P. E. (2009). Convectively coupled equatorial waves. *Reviews of Geophysics*, 47(2).

Knippertz, P., Gehne, M., Kiladis, G., Kikuchi, K., Rasheeda Satheesh, A., Roundy, P., Yang, G., Žagar, N., Dias, J., Fink, A., Methven, J., Schlueter, A., Sielmann, F., and Wheeler, M. (2022). The Intricacies of Identifying Equatorial Waves. *Quarterly Journal of the Royal Meteorological Society*, 148:2814–2852.

Kummerow, C., Hong, Y., Olson, W., Yang, S., Adler, R., Mccollum, J., Ferraro, R., Petty, G., Shin, D.-B., and Wilheit, T. (2001). The Evolution of the Goddard Profiling Algorithm (GPROF) for Rainfall Estimation From Passive Microwave Sensors. *Journal of Applied Meteorology*, 40:1801–1820.

Kummerow, C., Olson, W., and Giglio, L. (1996). A simplified scheme for obtaining precipitation and vertical hydrometeor profiles from passive microwave sensors. *IEEE Transactions on Geoscience and Remote Sensing*, 34(5):1213–1232.

Lakew, H., Moges, S., Dereje, H., and Asfaw, D. (2020). Hydrological performance evaluation of multiple satellite precipitation products in the upper Blue Nile basin, Ethiopia. *Journal of Hydrology: Regional Studies*, 27:100664.

Legates, D. and Mccabe, G. (1999). Evaluating the Use Of "Goodness-of-Fit" Measures in Hydrologic and Hydroclimatic Model Validation. *Water Resources Research*, 35:233–241.

Lemos, I., Lima, A., and Duarte, M. (2020). thresholdmodeling: A python package for modeling excesses over a threshold using the peak-over-threshold method and the generalized pareto distribution. *Journal of Open Source Software*, 5:2013.

Li, S. and Robertson, A. W. (2015). Evaluation of Submonthly Precipitation Forecast Skill from Global Ensemble Prediction Systems. *Mon. Wea. Rev.*, 143:2871–2889.

Lima, C., Lall, U., Jebara, T., and Barnston, A. (2009). Statistical Prediction of ENSO from Subsurface Sea Temperature Using a Nonlinear Dimensionality Reduction. *Journal of Climate*, 22:4501–4519.

Liu, J., Tang, Y., Wu, Y., Li, T., Wang, Q., and Chen, D. (2021). Forecasting the Indian Ocean Dipole With Deep Learning Techniques. *Geophysical Research Letters*, 48(20):e2021GL094407.

MacLeod, D., Graham, R., O'Reilly, C., Otieno, G., and Todd, M. (2021). Causal pathways linking different flavours of ENSO with the Greater Horn of Africa short rains. *Atmospheric Science Letters*, 22(2):e1015.

Macleod, D. A., Dankers, R., Graham, R., Guigma, K., Jenkins, L., Todd, M. C., Kiptum, A., Kilavi, M., Njogu, A., and Mwangi, E. (2021). Drivers and subseasonal predictability of heavy rainfall in Equatorial East Africa and relationship with flood risk. *J. Hydrometeor.*, 22:887–903.

Madden, R. A. and Julian, P. R. (1971). Detection of a 40–50 Day Oscillation in the Zonal Wind in the Tropical Pacific. *Journal of Atmospheric Sciences*, 28(5):702 – 708.

Madden, R. A. and Julian, P. R. (1972). Description of Global-Scale Circulation Cells in the Tropics with a 40–50 Day Period. *Journal of Atmospheric Sciences*, 29(6):1109 – 1123.

Maidment, R., Grimes, D., Allan, R., Greatrex, H., Rojas, O., and Leo, O. (2013). Evaluation of satellite-based and model re-analysis rainfall estimates for Uganda. *Meteorological Applications*, 20:308–317.

Maier-Gerber, M., Fink, A. H., Riemer, M., Schoemer, E., Fischer, C., and Schulz, B. (2021). Statistical-Dynamical Forecasting of Subseasonal North Atlantic Tropical Cyclone Occurrence. *Wea. Forecasting*, 36:2127–2142.

Maranan, M., Fink, A., Knippertz, P., Amekudzi, L., Atiah, W., and 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*, 21:729–749.

Marchant, R., Mumbi, C., Behera, S., and Yamagata, T. (2007). The Indian Ocean Dipole – the unsung driver of climate variability in East Africa. *African Journal of Ecology*, 45:4 – 16.

Marsham, J. H., Dickson, N. S., Garcia-Carreras, L., Lister, G. M. S., Parker, D. J., Knippertz, P., and Birch, C. E. (2013). The role of moist convection in the West African monsoon system: Insights from continental-scale convection-permitting simulations. *Geophys. Res. Lett.*, 40:1843–1849.

Žagar, N. (2017). A global perspective of the limits of prediction skill of NWP models. *Tellus*, 16A:1317573.

Matsuno, T. (1966). Quasi-Geostrophic Motions in the Equatorial Area. *Journal of the Meteorological Society of Japan. Ser. II*, 44(1):25–43.

Mccollum, J., Gruber, A., and Ba, M. (2000). Discrepancy between Gauges and Satellite Estimates of Rainfall in Equatorial Africa. *Journal of Applied Meteorology - J APPL METEOROL*, 39:666–679.

Mekonnen, A. and Thorncroft, C. D. (2016). On mechanisms that determine synoptic time scale convection over East Africa. *International Journal of Climatology*, 36(12):4045–4057.

Meyers, G., McIntosh, P., Pigot, L., and Pook, M. (2007). The Years of El Niño, La Niña, and Interactions with the Tropical Indian Ocean. *Journal of Climate - J CLIMATE*, 20:2872–2880.

Monsieurs, E., Kirschbaum, D., Jackson, T., Maki Mateso, J.-C., Jacobs, L., Plisnier, P.-D., Thiery, W., Augusta, U., Didace, M., Mugaruka Bibentyo, T., Bamulezi, G., Ilombe Mawe, G., Luc, B., Kankurize, C., Michellier, C., Thomas, S., Kervyn, F., Kervyn, M., Demoulin, A., and Dewitte, O. (2018). Evaluating TMPA Rainfall over the Sparsely Gauged East African Rift. *Journal of Hydrometeorology*, 19:1507–1528.

Morley, C. K., Ngenoh, D. K., and Ego, J. K. (1999). Introduction to the East African Rift System. pages 1–18. American Association of Petroleum Geologists.

Murphy, A. H. (1993). What Is a Good Forecast? An Essay on the Nature of Goodness in Weather Forecasting. *Weather and Forecasting*, 8(2):281 – 293.

Nash, J. and Sutcliffe, J. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 10(3):282–290.

Nicholson, S. (1996). A review of climate dynamics and climate variability in Eastern Africa (1996) Limnology, Climatology and paleoclimatology of the East African lakes. *edited by: Johnson, TC and Odada, E., Gordon & Breach, Amsterdam*, pages 25–56.

Nicholson, S. (2015). Long-term variability of the East African 'short rains' and its links to large-scale factors. *International Journal of Climatology*, 35.

Nicholson, S., Klotter, D., and Hartman, A. (2021). Lake-effect rains over Lake Victoria and their association with Mesoscale Convective Systems. *Journal of Hydrometeorology*, pages 1353–1368.

Nicholson, S. E. (2017). Climate and climatic variability of rainfall over eastern Africa. *Reviews of Geophysics*, 55(3):590–635.

Nicholson, S. E. (2018a). The ITCZ and the Seasonal Cycle over Equatorial Africa. *Bulletin of the American Meteorological Society*, 99(2).

Nicholson, S. E., Fink, A. H., and Funk, C. (2018b). Assessing recovery and change in West Africa's rainfall regime from a 161-year record. *International Journal of Climatology*, 38(10):3770–3786.

O, S. and Kirstetter, P.-E. (2018). Evaluation of diurnal variation of GPM IMERG-derived summerprecipitation over the contiguous US using MRMS data. *Quart. J. Roy. Meteor. Soc.*, 144:270–281.

OCHA (2020). Eastern Africa Region: Regional Floods and Locust Outbreak Snapshot January 2020. Accessed 31 March 2023, `https://reliefweb.int/report/south-sudan/eastern-africa-region-regional-floods-and-locust-outbreak-snapshot-january-2020`.

Oettli, P. and Camberlin, P. (2005). Influence of topography on monthly rainfall distribution over East Africa. *Clim. Res.*, 28:199–212.

Ogallo, L. J. (1989). The spatial and temporal patterns of the East African seasonal rainfall derived from principal component analysis. *International Journal of Climatology*, 9(2):145–167.

Ogallo, L. J., Janowiak, J., and Halpert, M. (1988). Teleconnection between Seasonal Rainfall over East Africa and Global Sea Surface Temperature Anomalies. *Journal of the Meteorological Society of Japan. Ser. II*, 66(6):807–822.

Ogwang, B. A., Chen, H., Li, X., and Gao, C. (2014). The influence of Topography on East African October to December climate: Sensitivity experiments with RegCM4. *Adv. Meteorol.*, 2014:14.

Palmer, P. I., Wainwright, C. M., Dong, B., Maidment, R. I., Wheeler, K. G., Gedney, N., Hickman, J. E., Madani, N., Folwell, S. S., Abdo, G., Allan, R. P., Black, E. C. L., Feng, L., Gudoshava, M., Haines, K., Huntingford, C., Kilavi, M., Lunt, M. F., Shaaban, A., and Turner, A. G. (2023). Drivers and impacts of Eastern African rainfall variability. *Nature Reviews Earth & Environment*, 4:254–270.

Parker, D. J., Blyth, A. M., Woolnough, S. J., Dougill, A. J., Bain, C. L., de Coning, E., Diop-Kane, M., dre Kamga Foamouhoue, A., Lamptey, B., Ndiaye, O., Ruti, P., Adefisan, E. A., Amekudzi, L. K., Antwi-Agyei, P., an d Carlo Cafaro, C. E. B., Carr, H., Chanzu, B., Clarke, S. J., Coskeran, H., Danuor, S. K., de Andrade, F. M., Diakaria, K., Dione, C., Diop, C. A., Fletcher, J. K., Gaye, A. T., Groves, J. L., Gudoshava, M., Hartley, A. J., Hirons, L. C., yaku Ibrahim, I., James, T. D., Lawal, K. A., Marsham, J. H., Mutemi, J. N., Okogbue, E. C., Olaniyan, E., Omotosho, J. B., Portuphy, J., Roberts, A. J., Schwendike, J., Segele, Z. T., Stein, T. H. M., Taylor, A. L., Taylor, C. M., rs, T. A. W., Webster, S., Woodhams, B. J., and Youds, L. (2022). The African SWIFT Project: Growing Science Capability to Bring about a Revolution in Weather Prediction. *Bulletin of the American Meteorological Society*, 103(2):E349 – E369.

Parker, D. J. and Diop-Kane, M., editors (2017). *Meteorology of West Africa: The Forecaster's Handbook*. John Wiley and Sons.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., and Louppe, G. (2012). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Peel, M., Finlayson, B., and Mcmahon, T. (2007). Updated World Map of the Koppen-Geiger Climate Classification. *Hydrology and Earth System Sciences Discussions*, 4:1633–1644.

Phillips, J. and Mcintyre, B. (2000). ENSO and interannual rainfall variability in Uganda: Implications for agricultural management. *International Journal of Climatology*, 20:171 – 182.

Pohl, B. and Camberlin, P. (2006a). Influence of the Madden-Julian Oscillation on East African rainfall. I: Intraseasonal variability and regional dependency. *Quart. J. Roy. Meteor. Soc.*, 132:2521–2539.

Pohl, B. and Camberlin, P. (2006b). Influence of the Madden–Julian Oscillation on East African rainfall: II. March–May season extremes and interannual variability. *Quart. J. Roy. Meteor. Soc.*, 132:2541–2558.

Richardson, L. F. (1922). *Weather prediction by numerical process*. Cambridge University Press.

Riley, E. M., Mapes, B. E., and Tulich, S. N. (2011). Clouds Associated with the Madden–Julian Oscillation: A New Perspective from CloudSat. *Journal of the Atmospheric Sciences*, 68(12):3032 – 3051.

Roundy, P. E. and Frank, W. M. (2004). A Climatology of Waves in the Equatorial Region. *Journal of the Atmospheric Sciences*, 61(17):2105 – 2132.

Saji, N. H., Goswami, B. N., Vinayachandran, P. N., and Yamagata, T. (1999). A dipole mode in the tropical Indian Ocean. *Nature*, 401:360–363.

Satheesh, R. A., Knippertz, P., Fink, H. A., Walz, E.-M., and Gneiting, T. (2023). Sources of predictability of synoptic-scale rainfall variability during the West African summer. *Quart. J. Roy. Meteor. Soc.*, pages 1–25.

Schlueter, A., Fink, A. H., and Knippertz, P. (2019b). A Systematic Comparison of Tropical Waves over Northern Africa. Part II: Dynamics and Thermodynamics. *Journal of Climate*, 32(9):2605 – 2625.

Schlueter, A., Fink, A. H., Knippertz, P., and Vogel, P. (2019c). A systematic comparison of tropical waves over northern africa. part i: Influence on rainfall. *Journal of Climate*, 32(5):1501 – 1523.

Schlüter, A. (2019a). *Tropical waves and rainfall over Africa: Variability, mechanisms and potential for forecasting*. PhD thesis, Karlsruher Institut für Technologie (KIT).

Schulz, B. and Lerch, S. (2022). Machine learning methods for postprocessing ensemble forecasts of wind gusts: A systematic comparison. *Mon. Wea. Rev.*, 150:235–257.

Seregina, L. S., Fink, A. H., van der Linden, R., Elagib, N. A., and Pinto, J. G. (2018). A new and flexible rainy season definition: Validation for Greater Horn of Africa and application to rainfall trends. *Int. J. Climatol.*, 39:989–1012.

Shaaban, A. and Roundy, P. (2017). OLR perspective on the Indian Ocean Dipole with application to East African precipitation: OLR Perspective on the IOD. *Quarterly Journal of the Royal Meteorological Society*, 143:828–1843.

Specq, D. and Batté, L. (2022). Do subseasonal forecasts take advantage of Madden-Julian oscillation windows of opportunity. *Atmos. Res.*, 23(4).

Stellingwerf, S., Riddle, E., Hopson, T. M., Knievel, J. C., Brown, B., and Gebremichael, M. (2021). Optimizing precipitation forecsts for hydrological catchment in Ethiopia using statistical bias correction and multi-modeling. *Earth and space science*, 8:27.

Stern, M. E. (1963). Trapping of Low Frequency Oscillations in an Equatorial "Boundary Layer". *Tellus*, 15(3):246–250.

Straub, K., Haertel, P., and Kiladis, G. (2010). An Analysis of Convectively Coupled Kelvin Waves in 20 WCRP CMIP3 Global Coupled Climate Models. *Journal of Climate - J CLIMATE*, 23:3031–3056.

Straub, K. and Kiladis, G. (2002). Observations of a Convectively Coupled Kelvin Wave in the Eastern Pacific ITCZ. *Journal of The Atmospheric Sciences - J ATMOS SCI*, 59:30–53.

Suhas, E. and Goswami, B. N. (2010). Loss of significance and multidecadal variability of the Madden–Julian oscillation. *J. Climate*, 23:3739–3751.

Takayabu, Y. (1994). Large-Scale Cloud Disturbances Associated with Equatorial Waves: Part II: Westward-Propagating Inertio-Gravity Waves. *Journal of the Meteorological Society of Japan. Ser. II*, 72:451–465.

Takayabu, Y. and Nitta, T. (1993). 3-5 Day-Period Disturbances Coupled with Convection over the Tropical Pacific Ocean. *Journal of the Meteorological Society of Japan*, 71:221–246.

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

Thiemig, V., Rojas, R., Zambrano-Bigiarini, M., Levizzani, V., and Roo, A. (2012). Validation of Satellite-Based Precipitation Products Over Sparsely-Gauged African River Basins. *Journal of Hydrometeorology*, 13:1760–1783.

Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika*, 18:267–276.

Tian, Y., Peters-Lidard, C., Eylander, J., Joyce, R., Huffman, G., Adler, R., Hsu, K., Turk, J., Garcia, M., and Zeng, J. (2009). Component Analysis of Errors in Satellite-Based Precipitation Estimates. *Journal of Geophysical Research*, 114:1–15.

Tompkins, A. M., Diongue-Niang, A., Parker, D. J., and Thorncroft, C. D. (2005). The African easterly jet in the ECMWF Integrated Forecast System: 4D-Var analysis. *Quarterly Journal of the Royal Meteorological Society*, 131(611):2861–2885.

Toreti, A., Bavera, D., Navarro, J. A., Cammalleri, C., de Jager, A., Ciollo, C. D., Hrast, A. E., Maetens, W., Magni, D., Masante, D., Mazzeschi, M., McCormick, N., Meroni, M., Rembold, F., and Spinoni, J. (2022). Drought in East Africa August 2022. Technical report.

UNEP (2006). Lake Victoria Basin Environment Outlook: Environment and Development. Accessed 21 May 2023, `https://wedocs.unep.org/bitstream/handle/20.500.11822/8613/LakeVictoria-Basin-Environment-Outlook.pdf`.

van der Linden, R., Fink, A. H., Pinto, J. G., Phan-Van, T., and Kiladis, G. N. (2016). Modulation of Daily Rainfall in Southern Vietnam by the Madden–Julian Oscillation and Convectively Coupled Equatorial Waves. *Journal of Climate*, 29(16):5801 – 5820.

Vannitsem, S., Wilks, D. S., and Messner, J. W. (2018). *Statistical Postprocessing of Ensemble Forecasts*. Elsevier.

Vashisht, A. and Zaitchik, B. (2022). Modulation of East African Boreal Fall Rainfall: Combined Effects of the Madden–Julian Oscillation (MJO) and El Niño–Southern Oscillation (ENSO). *Journal of Climate*, 35(6):2019 – 2034.

Vellinga, M. and Milton, S. (2018). Drivers of interannual variability of the East African 'Long Rains'. *Quarterly Journal of the Royal Meteorological Society*, 144.

Vigaud, N., Tippett, M. K., and Robertson, A. W. (2018). Probabilistic Skill of Subseasonal Precipitation Forecasts for the East Africa–West Asia Sector during September–May. *Weather and Forecasting*, 33(6):1513 – 1532.

Vitart, F. (2017). Madden-Julian Oscillation and teleconnections in the S2S database. *Quart. J. Roy. Meteor. Soc.*, 143:2210–2220.

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., Déqué, M., Ferranti, L., Fucile, E., Fuentes, M., Hendon, H., Hodgson, J., Kang, H., Kumar, A., Lin, H., Liu, G., Liu, X., Malguzzi, P., Mallas, I., and Zhang, L. (2017). The subseasonal to Seasonal prediction (S2S) project database. *Bull. Amer. Meteor. Soc.*, 98:163–173.

Vogel, P., Knippertz, P., Fink, A. H., Schlueter, A., and Gneiting, T. (2018). Skill of global raw and postprocessed ensemble predictions of rainfall over Northern Tropical Africa. *Wea. Forecating*, 33:369–388.

Vogel, P., Knippertz, P., Fink, A. H., Schlueter, A., and Gneiting, T. (2020). Skill of global raw and postprocessed ensemble predictions of rainfall in the tropics. *Wea. Forecating*, 35:2367–2385.

Vogel, P., Knippertz, P., Gneiting, T., Fink, A. H., Klar, M., and Schlueter, A. (2021). Statistical Forecasts for the Occurrence of Precipitation Outperform Global Models over Northern Tropical Africa. *Geophysical Research Letters*, 48(3):e2020GL091022.

Wainwright, C. M., Finney, D. L., Kilavi, M., Black, E., and Marsham, J. H. (2021). Extreme rainfall in East Africa, October 2019–January 2020 and context under future climate change. *Weather*, 76(1):26–31.

Walker, D., Birch, C., Marsham, J., Scaife, A., Graham, R., and Segele, Z. (2019). Skill of dynamical and GHACOF consensus seasonal forecasts of East African rainfall. *Climate Dynamics*, 53:4911–4935.

Walker, G. T. (1925). CORRELATION IN SEASONAL VARIATIONS OF WEATHER—A FURTHER STUDY OF WORLD WEATHER. *Monthly Weather Review*, 53(6):252 – 254.

Walz, E., Maranan, M., van der Linden, R., Dink, A. H., and Knippertz, P. (2021). An IMERG-based optimal extended probabilistic climatology (EPC) as a benchmark

ensemble forecast for precipitation in the tropics ad subtropics. *Wea. Forecating*, 36:1561–1573.

Wang, X., Slawinska, J., and Giannakis, D. (2020). Extended-range statistical ENSO prediction through operator-theoretic techniques for nonlinear dynamics. *Scientific Reports*, 10.

Webster, P. J., Moore, A. M., Loschnigg, J. P., and Leben, R. R. (1999). Coupled ocean–atmosphere dynamics in the Indian Ocean during 1997–98. *Nature*, 401:356–360.

Wheeler, M. and Kiladis, G. N. (1999). Convectively Coupled Equatorial Waves: Analysis of Clouds and Temperature in the Wavenumber–Frequency Domain. *Journal of the Atmospheric Sciences*, 56(3):374 – 399.

Wheeler, M. C. and Hendon, H. H. (2004). An All-Season Real-Time Multivariate MJO Index: Development of an Index for Monitoring and Prediction. *Monthly Weather Review*, 132(8):1917 – 1932.

Wilks, D. S., editor (2011). *Statistical Methods in the Atmospheric Sciences*, volume 100 of *International Geophysics series*. Academic Press, 3 edition.

Wilks, D. S. (2016). "The stippling shows statistically significant grid points" how research results are routinely overstated and overinterpreted, and what to do about it. *Bull. Amer. Meteor. Soc.*, 97:2263–2273.

Williams, A. P. and Funk, C. (2011). A westward extension of the warm pool leads to a westward extension of the Walker circulation, drying eastern Africa. *Climate Dyn.*, 37:2417–2435.

WMO (2023). Boosting the provision of essential weather and climate data. Accessed 14 June 2023, `https://public.wmo.int/en/media/press-release/boosting-provision-of-essential-weather-and-climate-data`.

Woodhams, B. J., Birch, C. E., Marsham, J. H., Bain, C. L., Roberts, N. M., and Boyd, D. F. A. (2018). What is the added value of a convection-permitting model for forecasting extreme rainfall over tropical East Africa. *Mon. Wea. Rev.*, 146:2757–2780.

Wu, R., Kirtman, B., and Dool, H. (2009). An Analysis of ENSO Prediction Skill in the CFS Retrospective Forecasts. *Journal of Climate - J CLIMATE*, 22:1801–1818.

Yamagata, T., Behera, S., Luo, J.-J., Masson, S., Jury, M., and Rao, S. (2004). Coupled Ocean-Atmosphere Variability in the Tropical Indian Ocean. *Earth's Climate: The Ocean- Atmosphere Interaction. Geophysics Monograph Series*, 147.

Yanai, M. and Maruyama, T. (1966). Stratospheric Wave Disturbances Propagating over the Equatorial Pacific. *Journal of the Meteorological Society of Japan*, 44:291–294.

Yang, G.-Y., Ferrett, S., Woolnough, S., Methven, J., and Holloway, C. (2021). Real-Time Identification of Equatorial Waves and Evaluation of Waves in Global Forecasts. *Weather and Forecasting*, 36(1):171 – 193.

Yang, W., Seager, R., Cane, M. A., and Lyon, B. (2015). The Annual Cycle of East African Precipitation. *Journal of Climate*, 28(6):2385 – 2404.

Yasunaga, K. and Mapes, B. (2012). Differences between More Divergent and More Rotational Types of Convectively Coupled Equatorial Waves. Part I: Space–Time Spectral Analyses. *Journal of the Atmospheric Sciences*, 69(1):3 – 16.

Youds, L., Parker, D., Adefisan, E., Antwi-Agyei, P., Bain, C., Black, E., Blyth, A., Dougill, A., Hirons, L., Indasi, V., Lamptey, B., Marshall, F., Marsham, J., Stein, T., Taylor, C., Todd, M., Visman, E., and Woolnough, S. (2021). GCRF African SWIFT and ForPAc SHEAR White Paper on the Potential of Operational Weather Prediction to Save Lives and Improve Livelihoods and Economies in Sub-Saharan Africa. Report. © University of Leeds (2021). This work is licenced under Creative Commons (CC-BY 4.0).

Young, M. P., Williams, C. J. R., Chiu, J. C., Maidment, R. I., and Chen, S.-H. (2014). Investigation of Discrepancies in Satellite Rainfall Estimates over Ethiopia. *Journal of Hydrometeorology*, 15(6):2347 – 2369.

Zhang, C. (2005). Madden-Julian Oscillation. *Reviews of Geophysics*, 43(2).

Zhang, Y., Moges, S., and Block, P. (2016). Optimal Cluster Analysis for Objective Regionalization of Seasonal Precipitation in Regions of High Spatial-Temporal Variability: Application to Western Ethiopia. *Journal of Climate*, 29:3697–3717.

Zhao, S., Stuecker, M., Jin, F.-F., Feng, J., Ren, H., Zhang, W., and Li, J. (2020). Improved Predictability of the Indian Ocean Dipole Using a Stochastic Dynamical Model Compared to the North American Multimodel Ensemble Forecast. *Weather and Forecasting*, 35:379–399.

Appendix: Figures

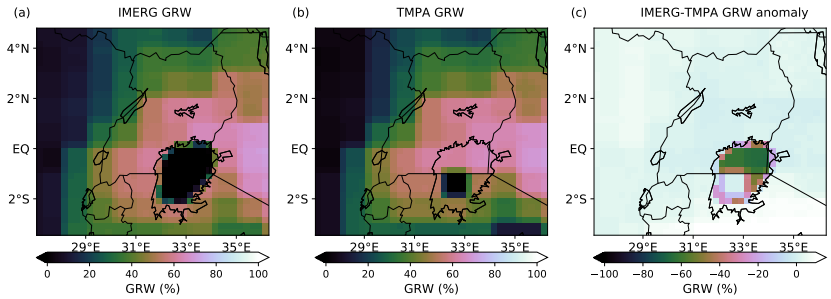


Figure A.1: The mean monthly gauge-relative weights (GRWs) (2001-2018) within; (a) IMERG and (b) TMPA. For both products the GRWs over almost all the pixel in Uganda but L. Victoria are the same. In IMERG over L. Victoria, the GRW follows the coastline, but this is not the case in TRMM. This difference over Lake Victoria can be clearly seen in the anomaly plot (c). This largely accounts for dipole seen over the lake in Fig. 3(d;e) in the manuscript.

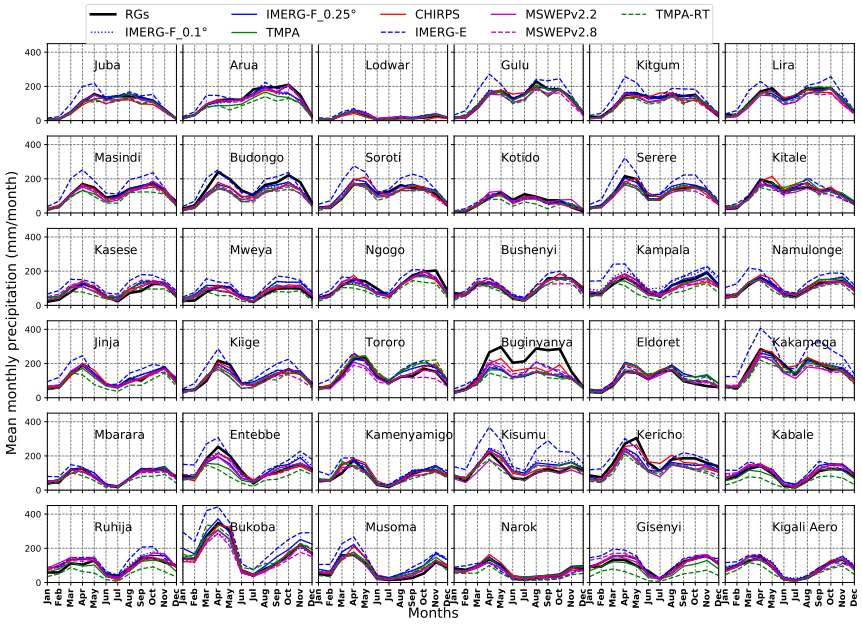


Figure A.2: Seasonal cycles of the 36 stations used in the study for the different datasets. In all the stations and for all the data sets, the seasonal cycles are generally well replicated. However, IMERG-E overestimates rainfall in almost all the stations, especially during the rainy seasons (i.e., MAM and SON). At Buginyanya station, all the SREs substantially underestimate monthly rainfall.

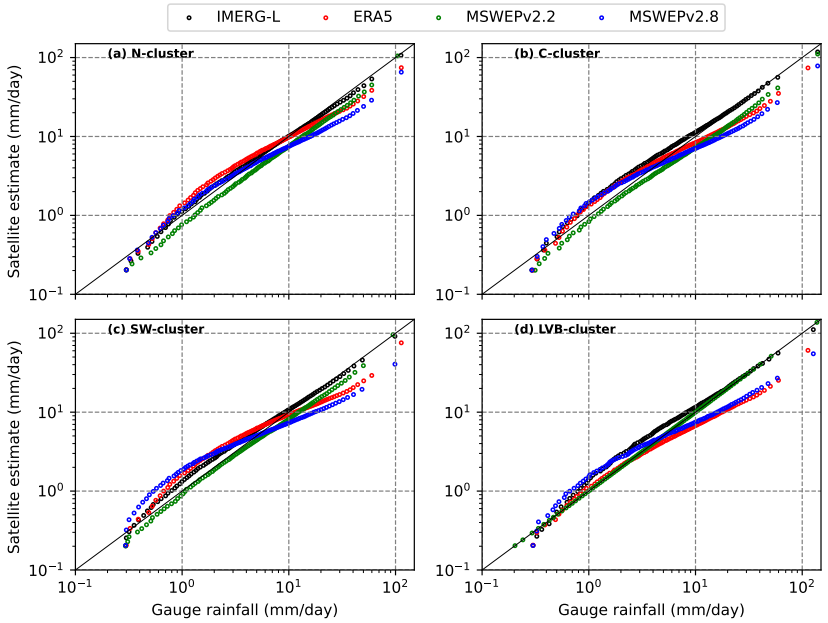


Figure A.3: Q-Q plots for MSWEPv2.2 and MSWEPv2.8 and the in-put datasets (IMERG-L and ERA5) used to create MSWEPv2.8 showing the distribution of the rain rates.

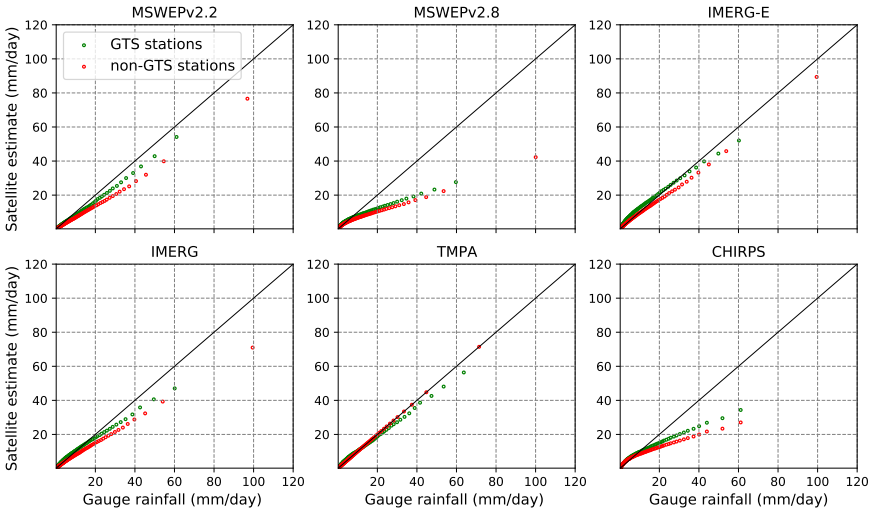


Figure A.4: Q-Q plots comparing the distribution of rain rates in GTS and non-GTS stations for each of the SRE. The plots shows that the distribution of rainfall rates in SREs differs from that of the RGs with the subsets of non-GTS stations differing more severely than the GTS stations.

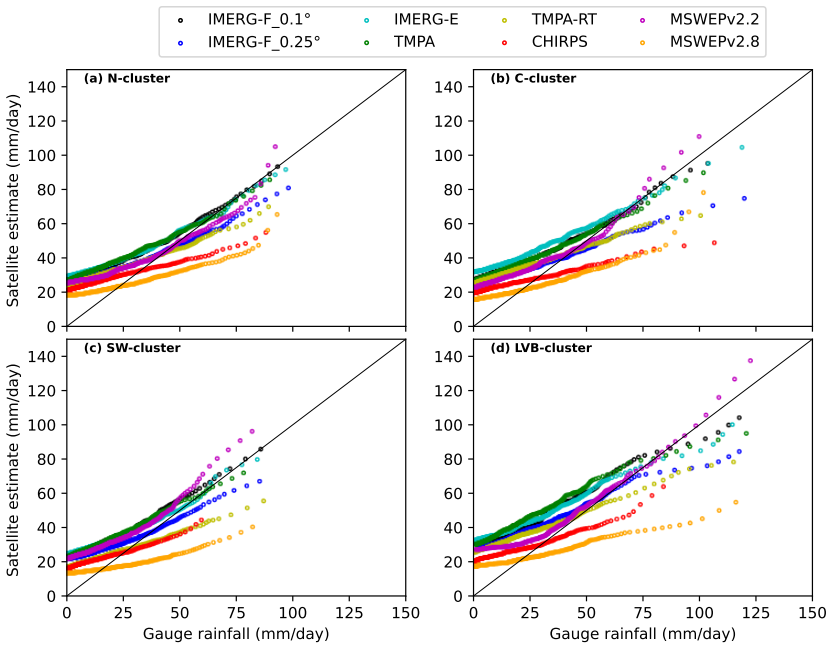


Figure A.5: QQ plots showing the distribution of extreme rainfall (rainfall amount greater than the 95th percentile of subset of rainy days subset) recorded by the SREs at the pixel nearest to the RGs.

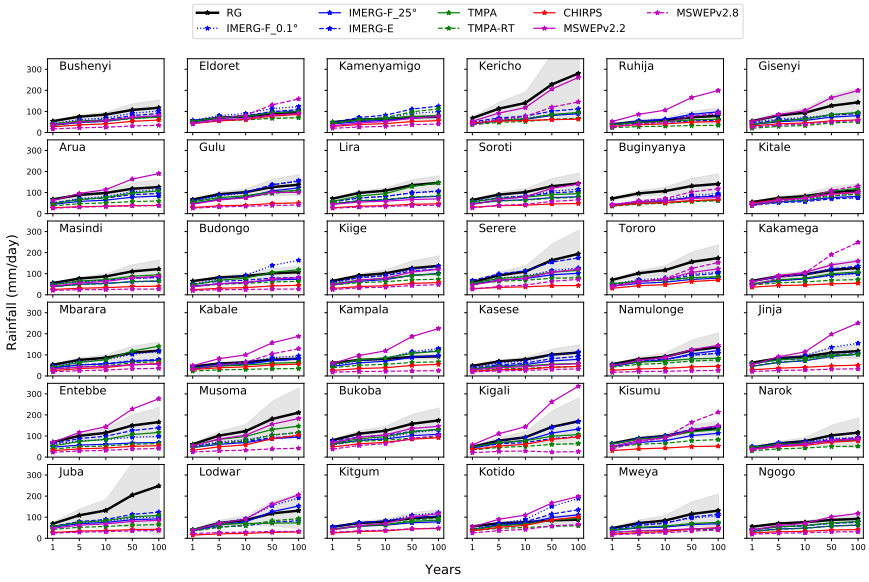


Figure A.6: Same as Fig. 4.12in the main text, but for all the stations.

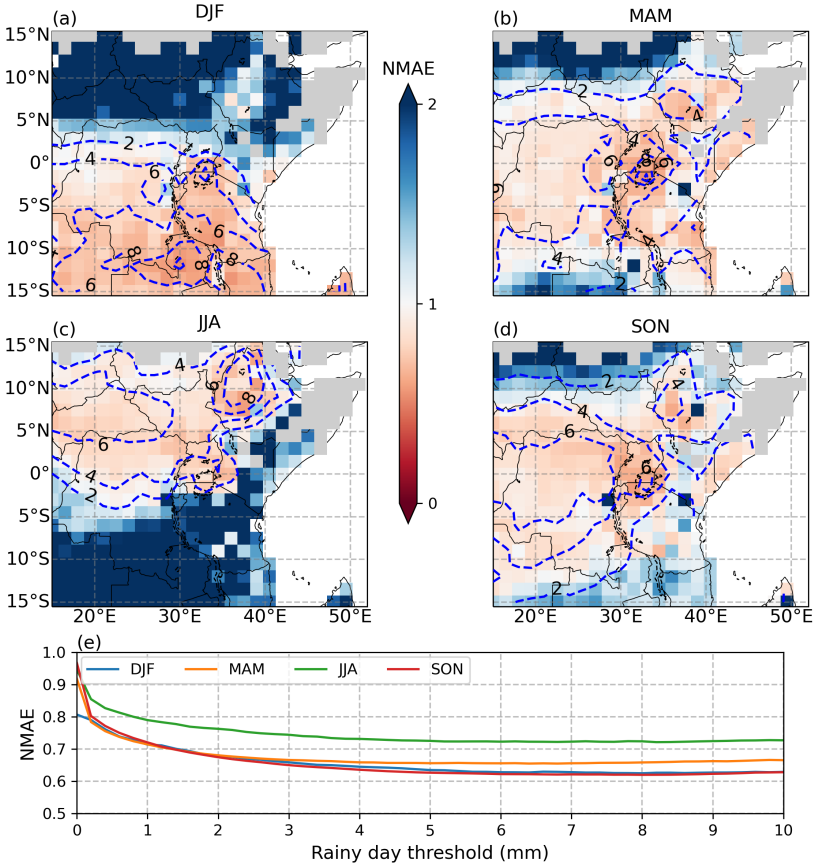


Figure A.7: Same as Fig. 5.1 in the main text but for normalised mean absolute error (NMAE).

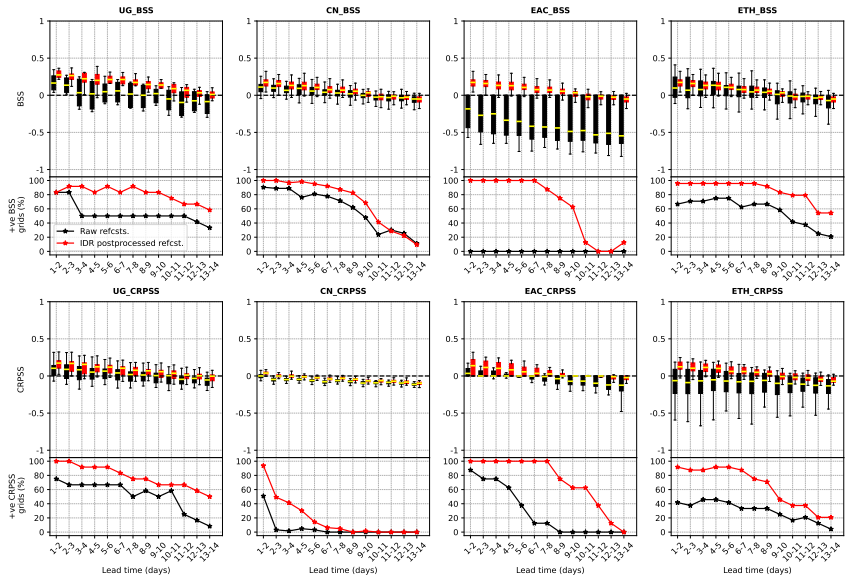


Figure A.8: Same as Fig. 5.7 in the main text but for 48-hour rainfall accumulation.

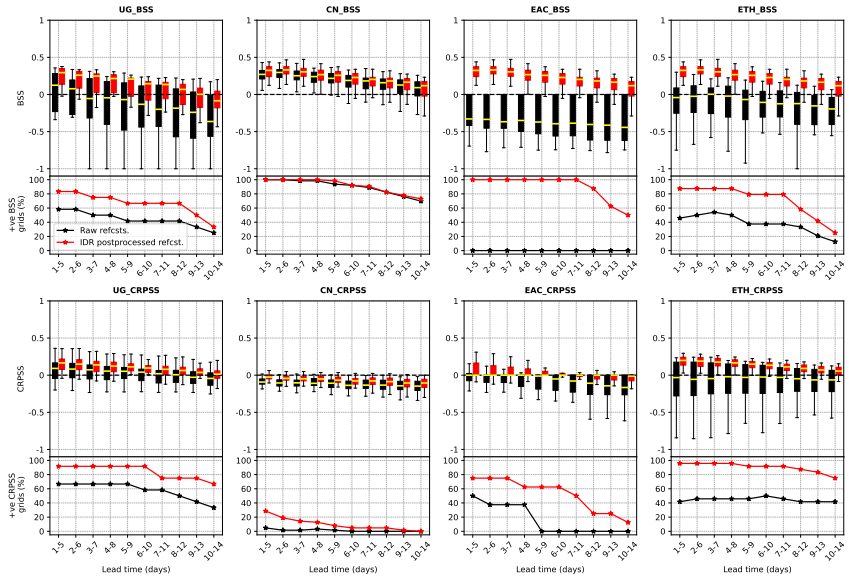


Figure A.9: Same as Fig. 5.7 in the main text but for 120-hour accumulation.

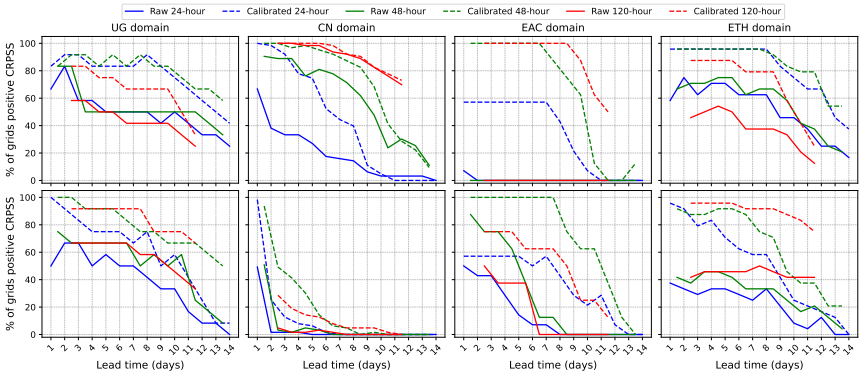


Figure A.10: Same as Fig. 5.8 in the main text but for the different regions.

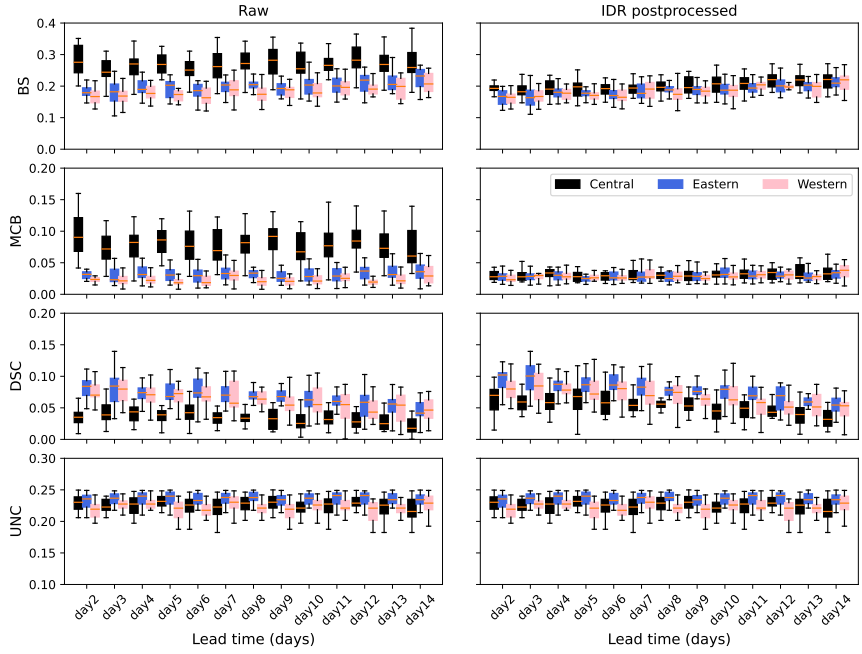


Figure A.11: CORP score decomposition for the grids used in the gauge analysis. The columns represent the raw and postprocessed versions of the reforecasts, while the rows are for the CORP score, miscalibration, discrimination and the uncertainty, respectively.

Acknowledgements

First and foremost, my greatest thanks go the Almighty God for creating me and taking care of me up to know. I owe all my accomplishments to you God!

All this would not be possible without the trust, guidance and patience of my Supervisors, Prof. Dr. Andreas Fink and Prof. Dr. Peter Knippertz. Many thanks you both! To Andreas, your confidence in me, right from the start, was a real source of motivation. I have learnt a lot from you, during our meetings and small talks. I am also grateful for extra the responsibilities you assigned me. I learnt a lot working as a Research Assistant in KASS-D. Moreover, I appreciated the extra financial security that the responsibility accorded me.

To my colleagues, I learnt a lot from each and everyone of you. Many thanks to my "third Supervisor", Marlon, for all the support. You made my PhD journey easier. Special thanks also go to Benedikt for you insightful inputs, especially on the statistical component to my project. Moreover, my appreciation also go Athul and Eva-Maria for sharing the code and for the discussions regarding my work. To Beth and Musa, I appreciate the discussions we had on East African weather, these helped shape parts of my work. My Corner Office-mates, Lea, Hyunju, Brett and Michael, thanks for creating a nice working environment. To the rest of group member, whom I won't mention by name so as not to writing a "500-page" thesis, you all made the Atmospheric Dynamics group a wonderful team to belong in.

I also thank the German government which sponsored my stay and studies at IMK-TRO, through DAAD program.

Greatest appreciation to my wife, Eva thank you for your Love and patience. I could always count on you and Crystal to keep me going, I Love you both. To my parents and sibling, although you were far, your belief in me, keeps me going, thank you.